

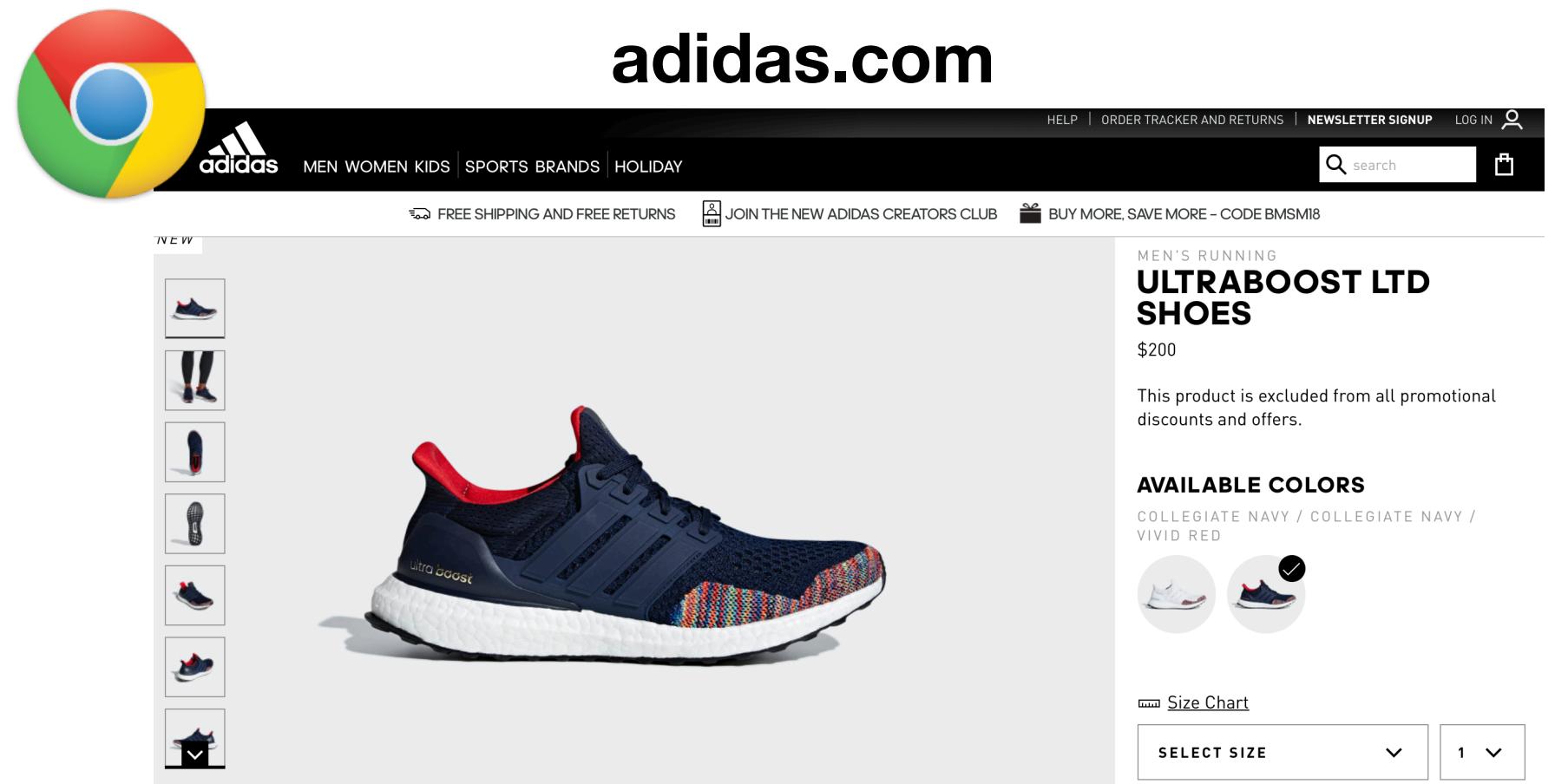
On the Privacy Implications of Real Time Bidding

Muhammad Ahmad Bashir
Thesis Proposal

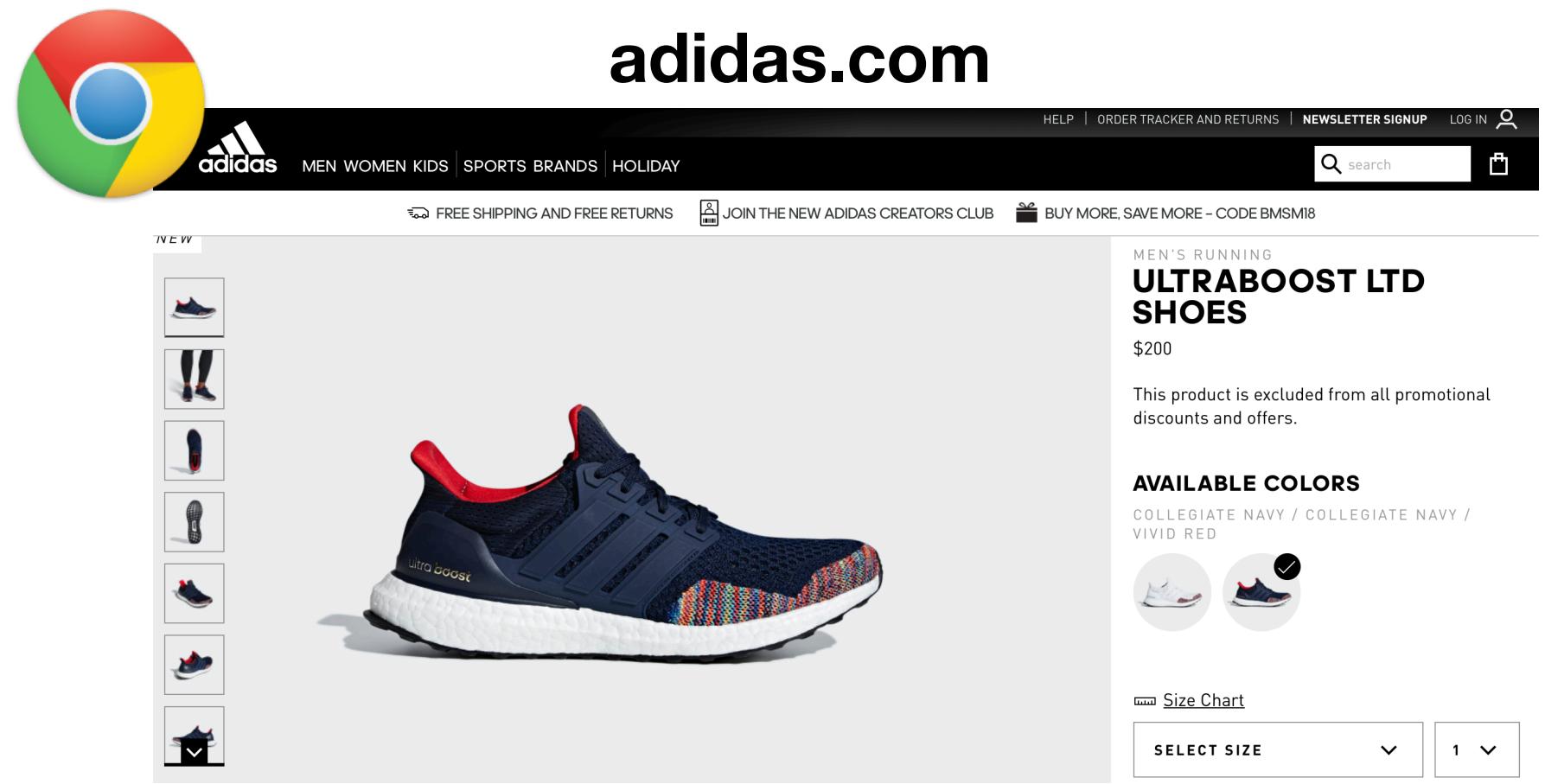


Northeastern University
College of Computer and Information Science

How Companies Track You on the Web

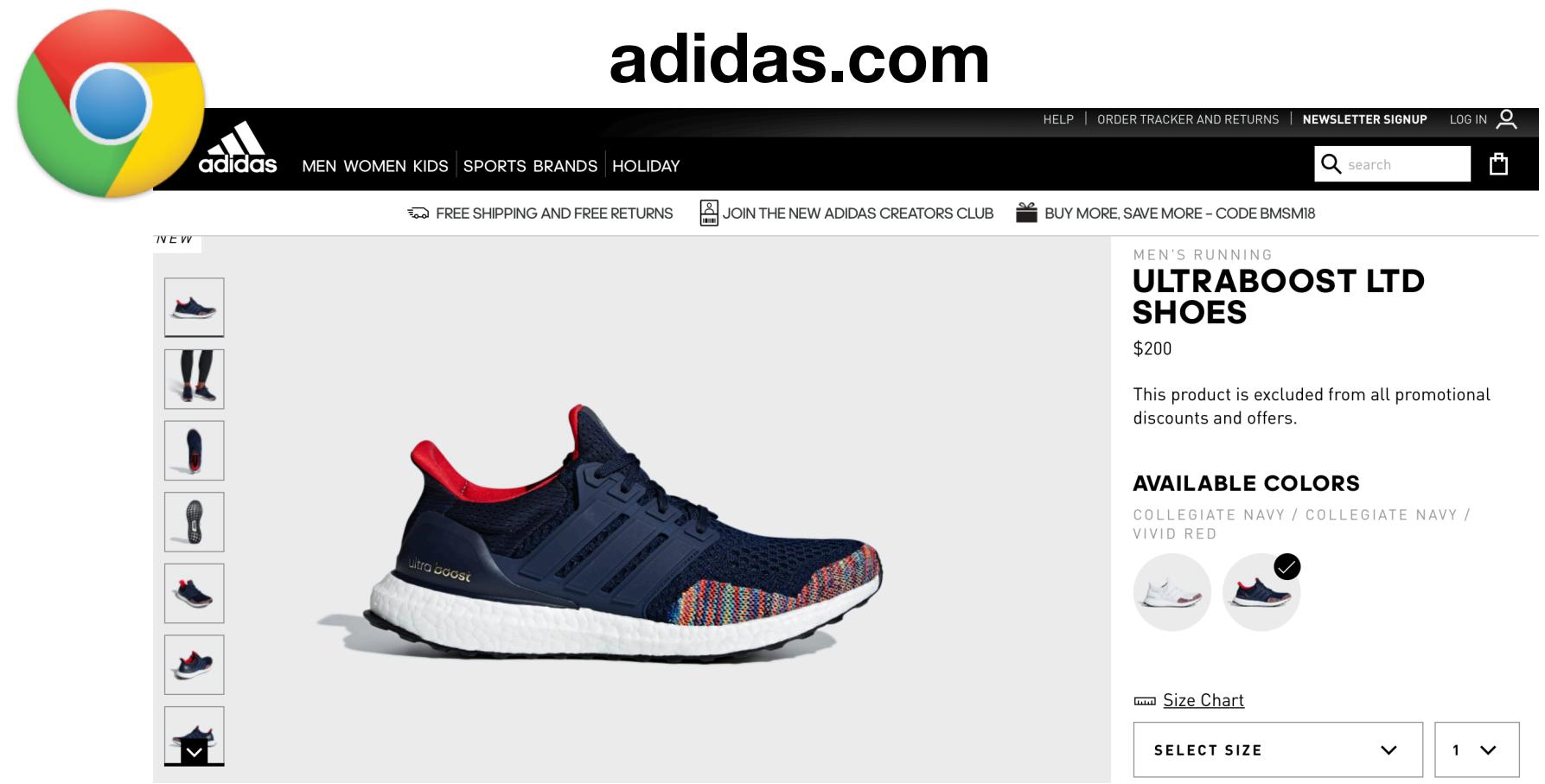


How Companies Track You on the Web



criteo.

How Companies Track You on the Web



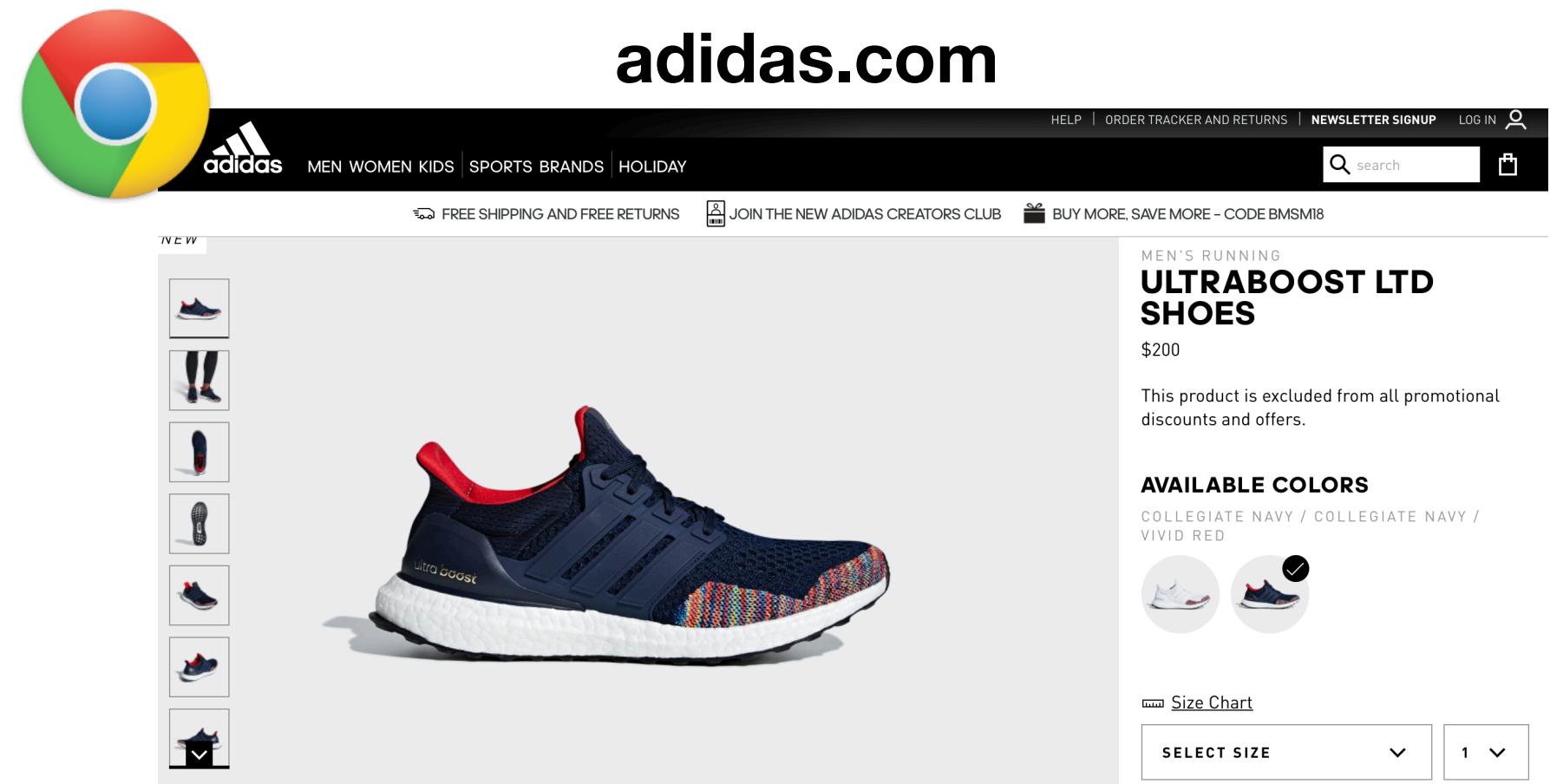
criteo.

User Profiles

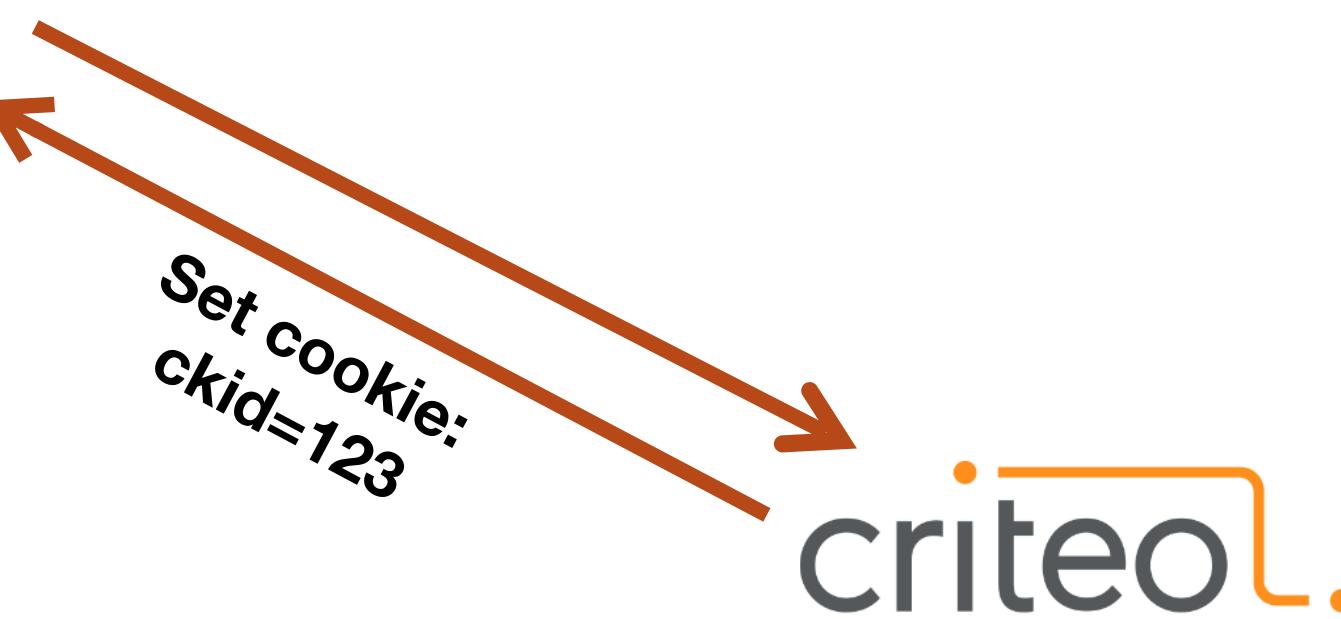
User A

- [babybump.com](#)
- [walmart, \(baby cribs\)](#)
- [\[redacted\]](#)
- [\[redacted\]](#)
- [\[redacted\]](#)

How Companies Track You on the Web



* Every domain has a different unique cookie for you

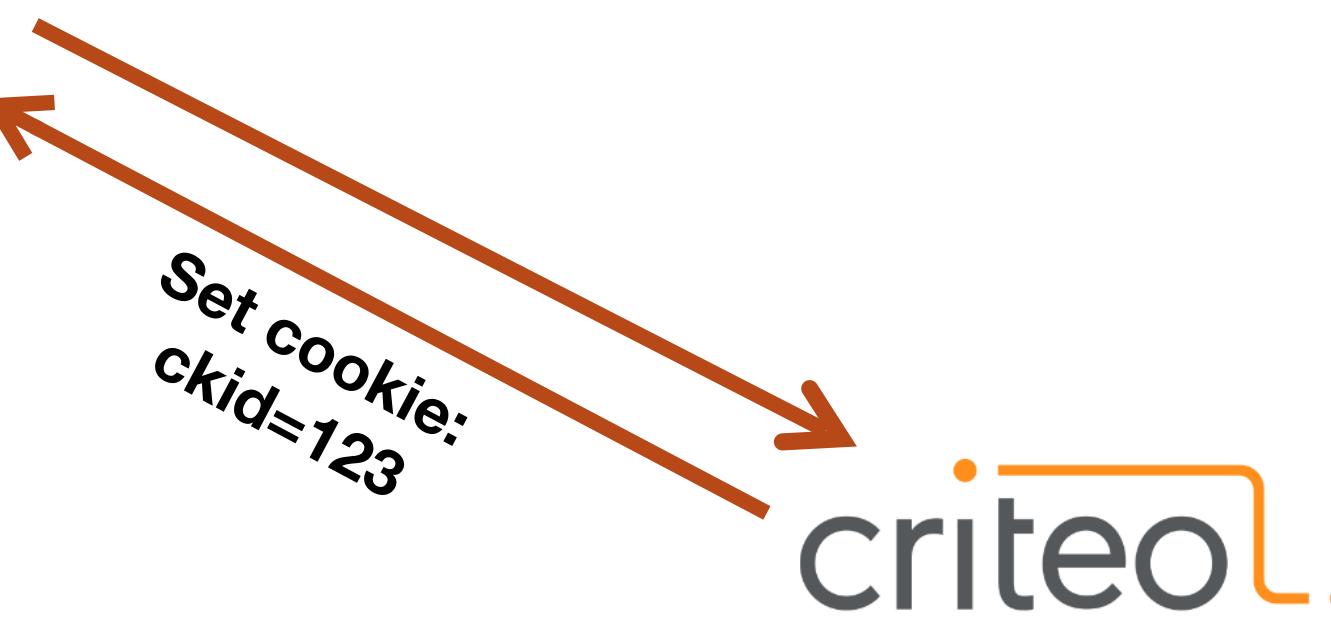
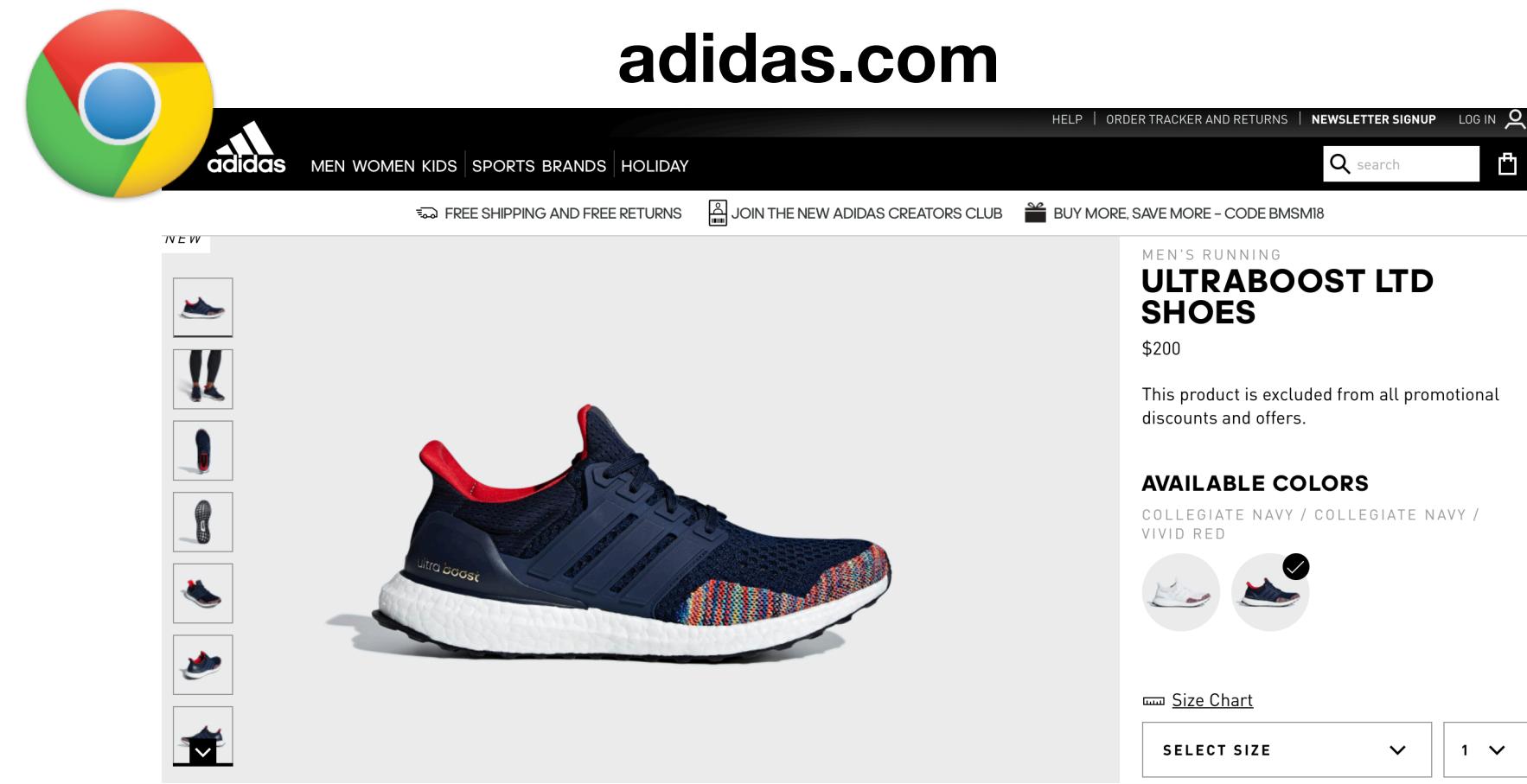


User Profiles

User A

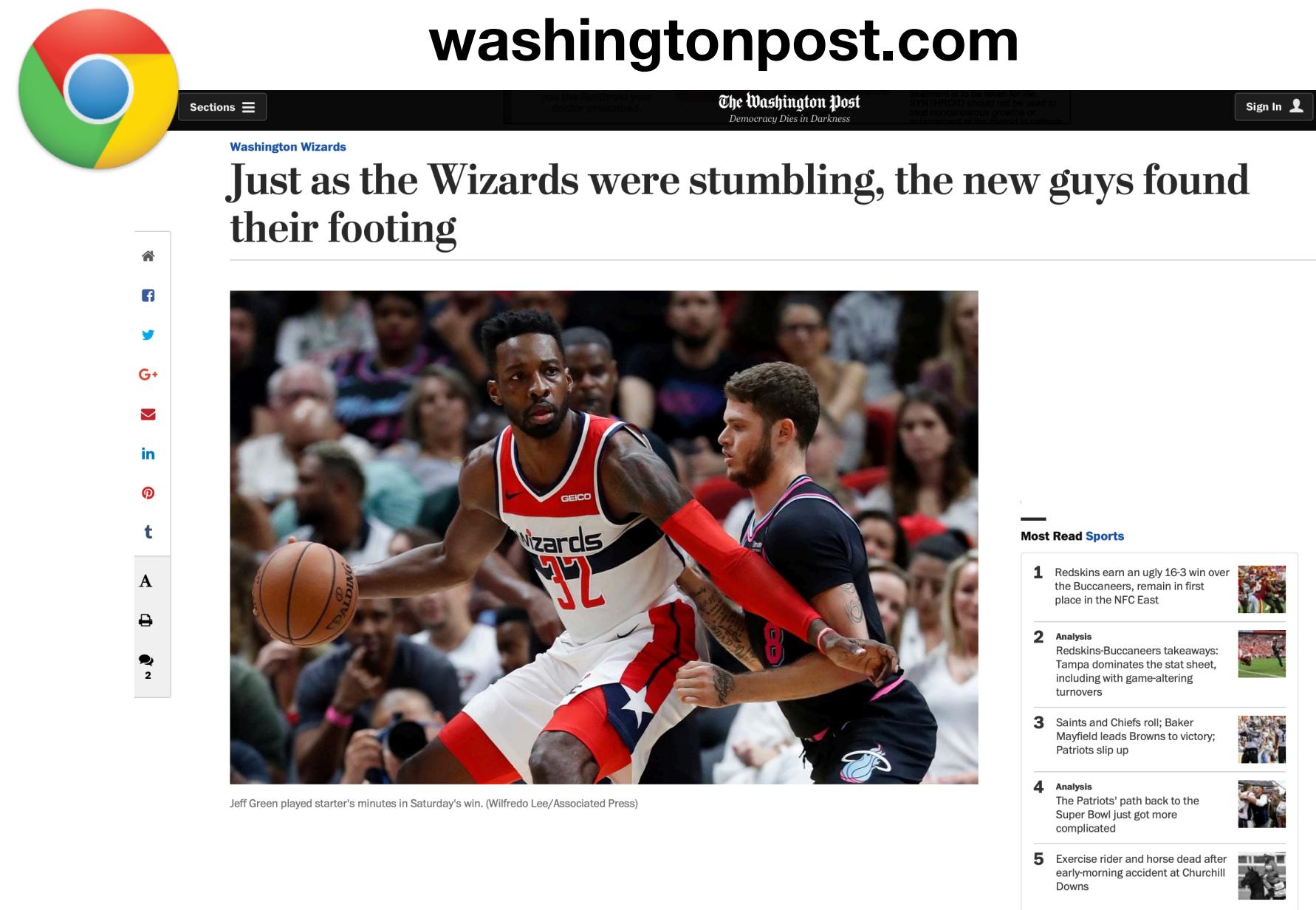
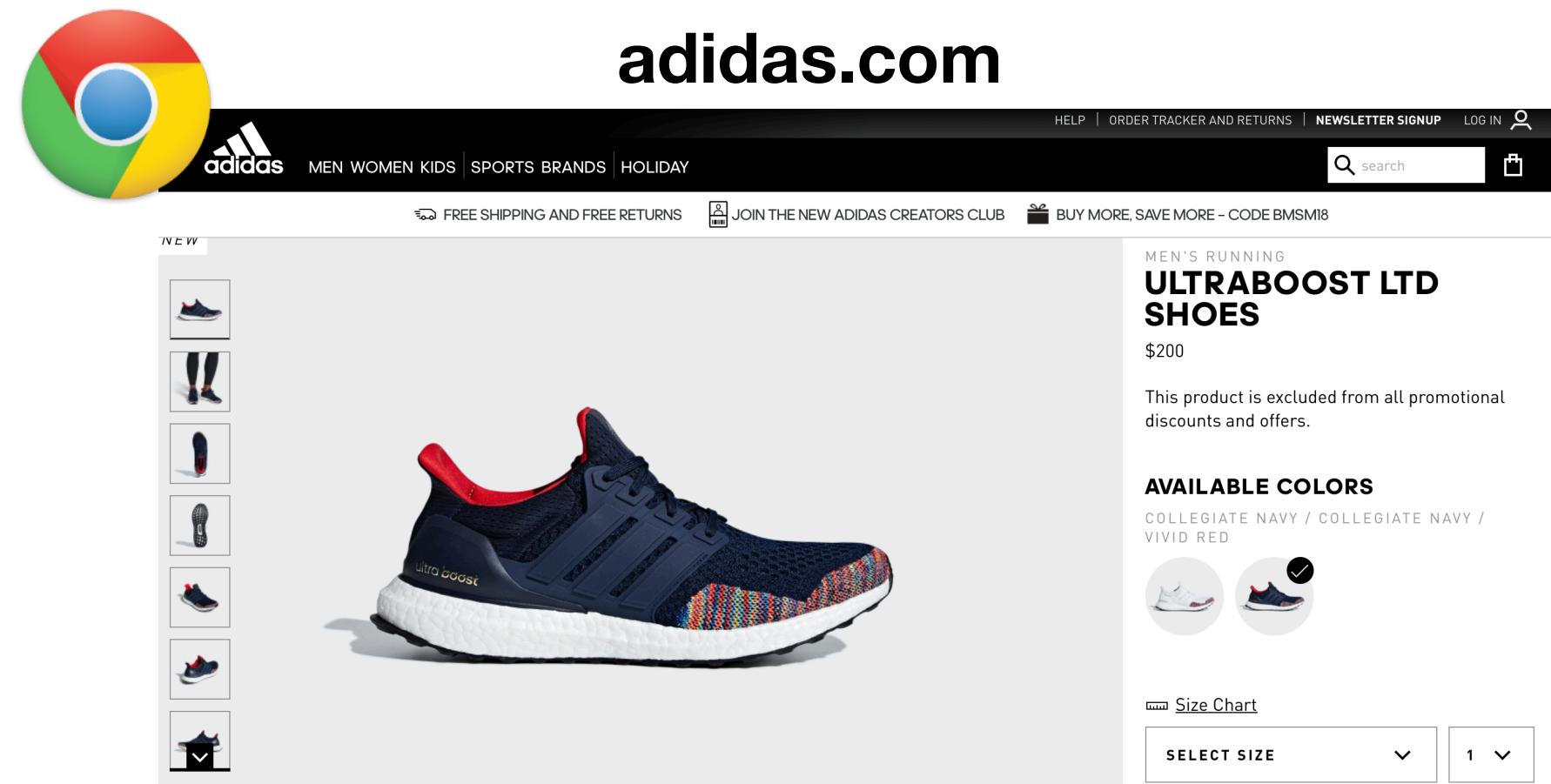
- babybump.com
- walmart, (baby cribs)
- -
- -
- -

How Companies Track You on the Web



* Every domain has a different unique cookie for you

How Companies Track You on the Web



* Every domain has a different unique cookie for you



User Profiles

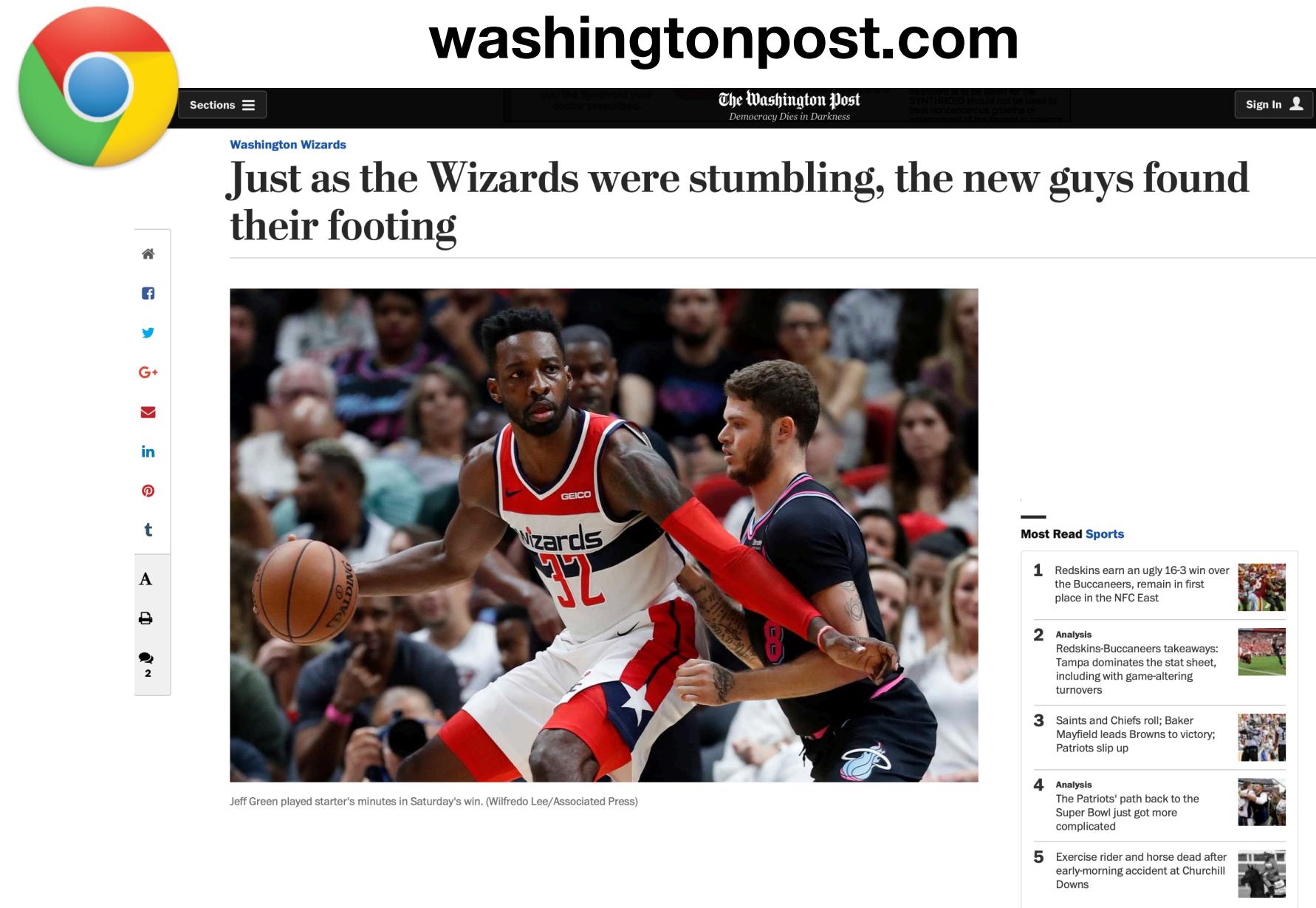
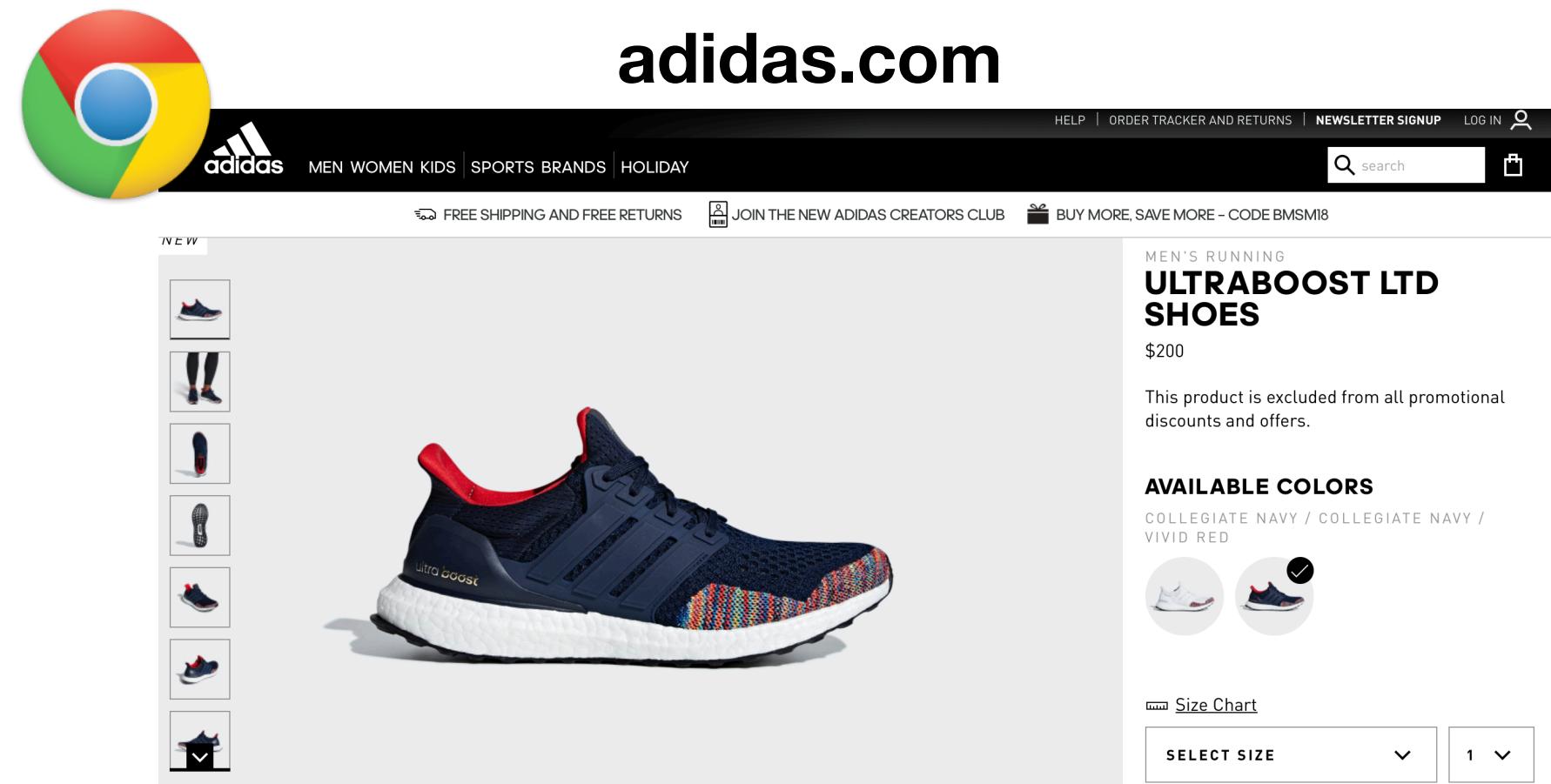
User A

- babybump.com
- walmart, (baby cribs)
-
-
-

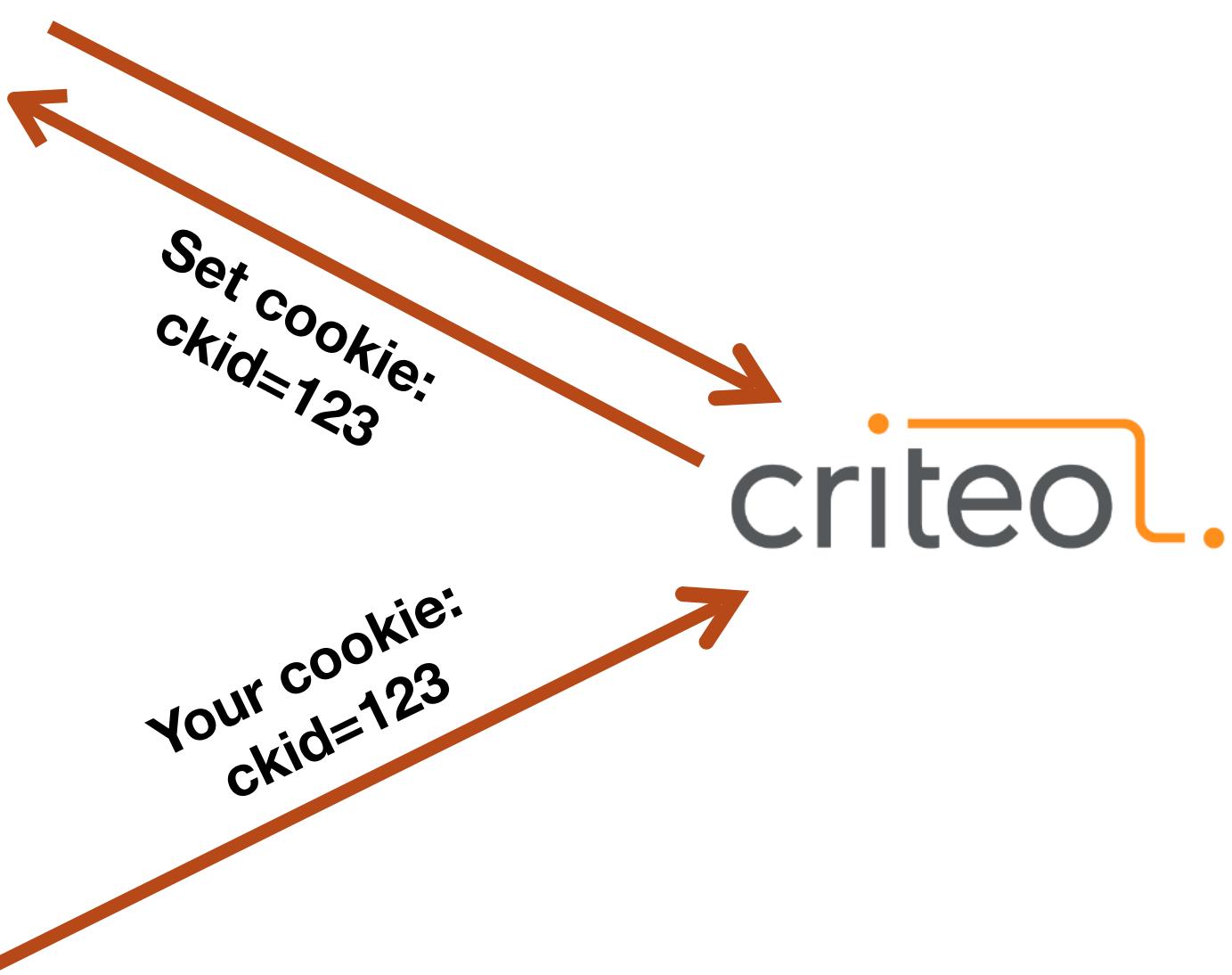
User 123

- adidas.com (ultraboost)

How Companies Track You on the Web



* Every domain has a different unique cookie for you



User Profiles

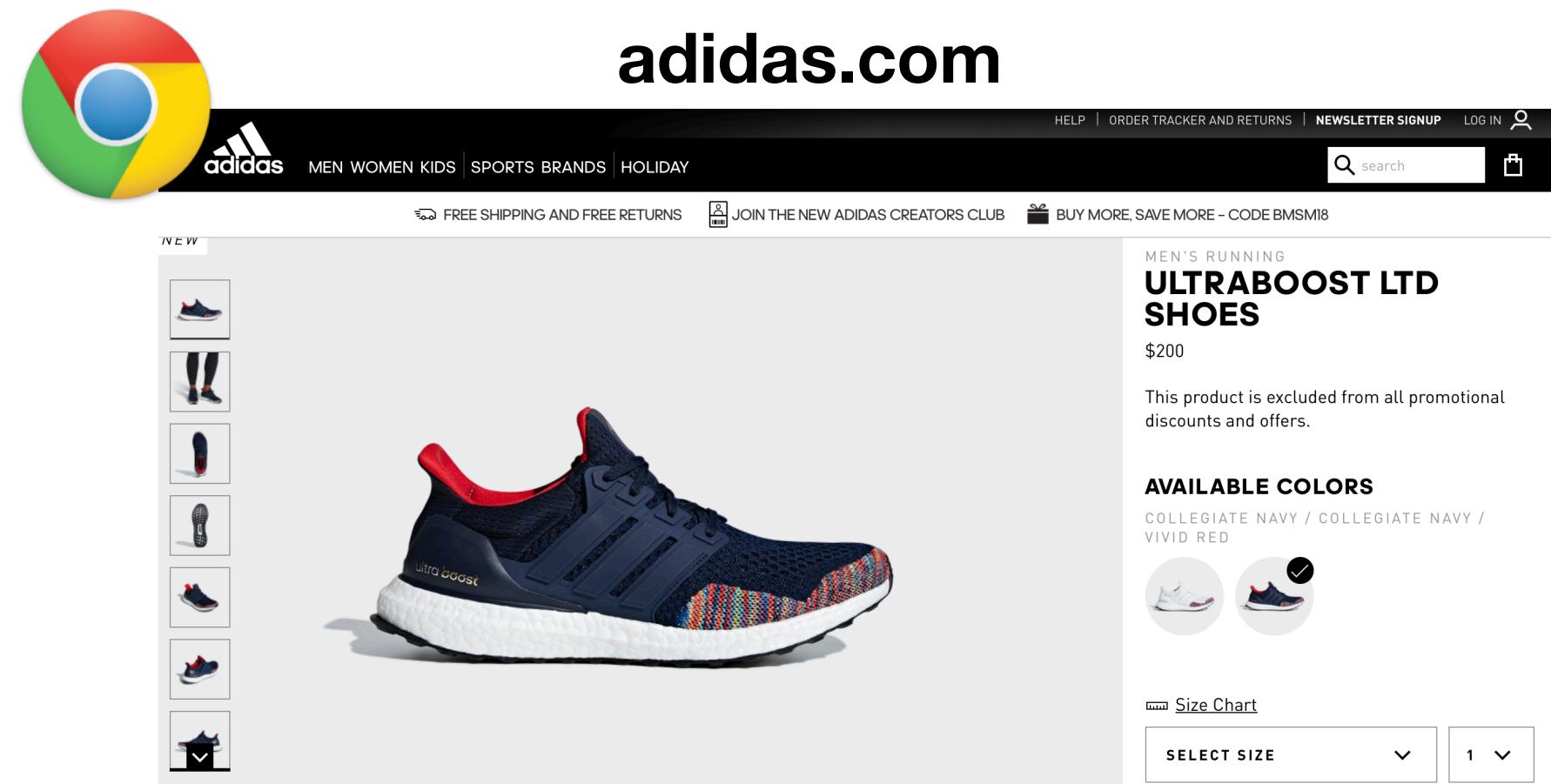
User A

- babybump.com
- walmart, (baby cribs)
-
-
-

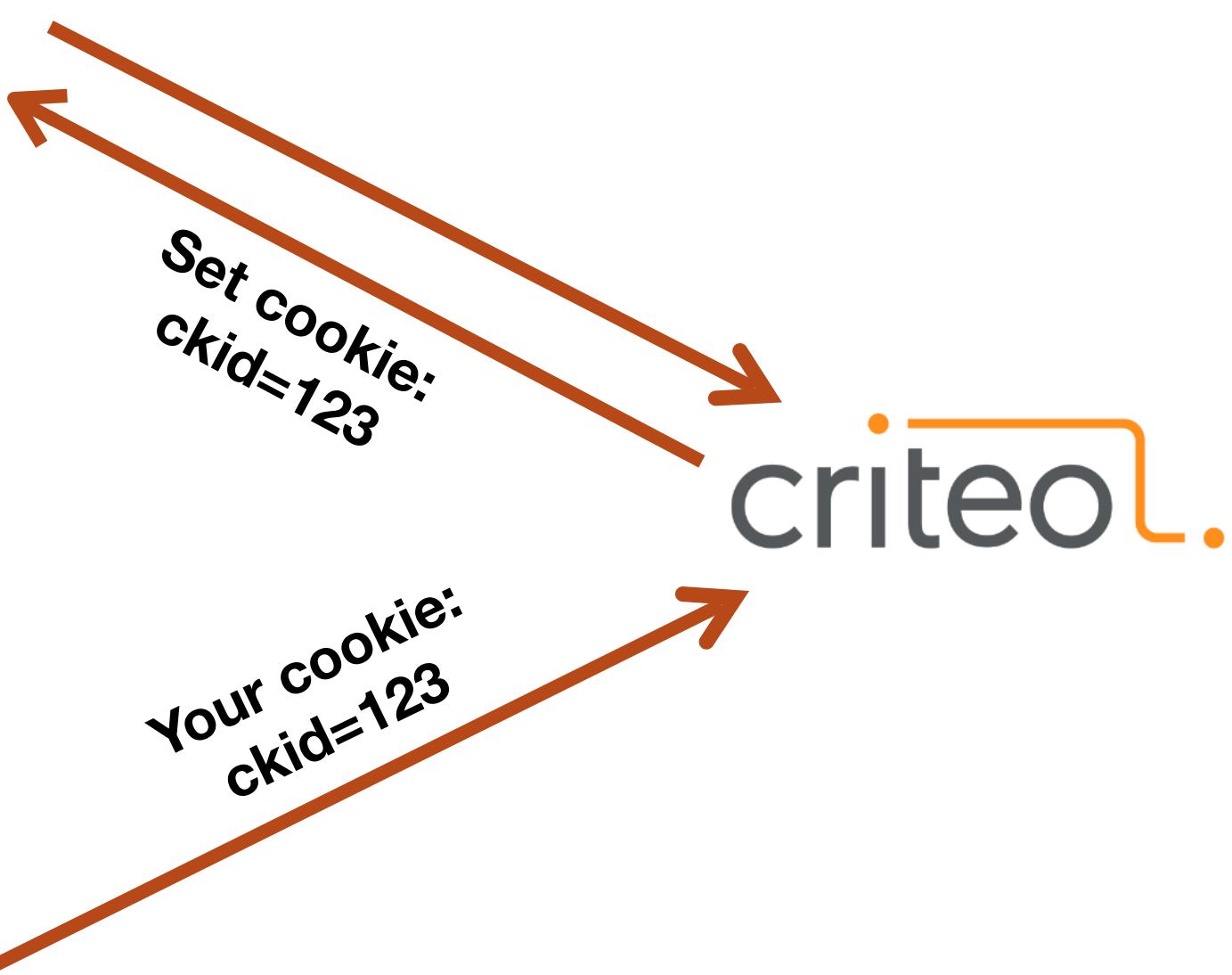
User 123

- adidas.com (ultraboost)

How Companies Track You on the Web



* Every domain has a different unique cookie for you



User Profiles

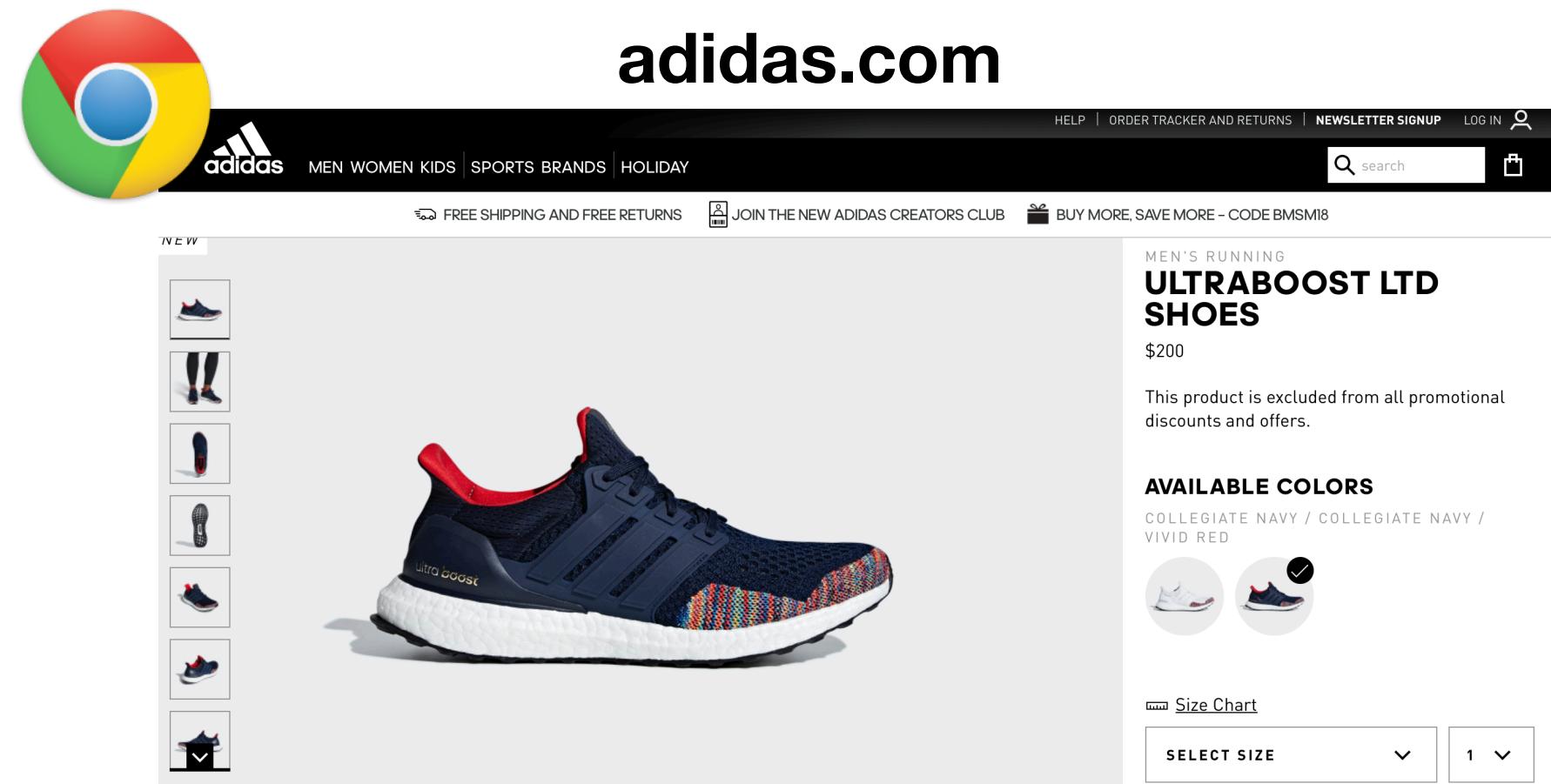
User A

- babybump.com
- walmart, (baby cribs)
- -
- -
- -

User 123

- adidas.com (ultraboost)
- washingtonpost (sports)

How Companies Track You on the Web



* Every domain has a different unique cookie for you



User Profiles

User A

- [babybump.com](#)
- [walmart](#), (baby cribs)
-
-
-

User 123

- [adidas.com](#) (ultraboost)
- [washingtonpost](#) (sports)

The Rise of Real Time Bidding (RTB)

Process of buying and selling ad impressions through real-time auctions.

The Rise of Real Time Bidding (RTB)

Process of buying and selling ad impressions through real-time auctions.

Publishers



Ad Networks



Advertisers



The Rise of Real Time Bidding (RTB)

Process of buying and selling ad impressions through real-time auctions.



The Rise of Real Time Bidding (RTB)

Process of buying and selling ad impressions through real-time auctions.



The Rise of Real Time Bidding (RTB)

Process of buying and selling ad impressions through real-time auctions.



The Rise of Real Time Bidding (RTB)

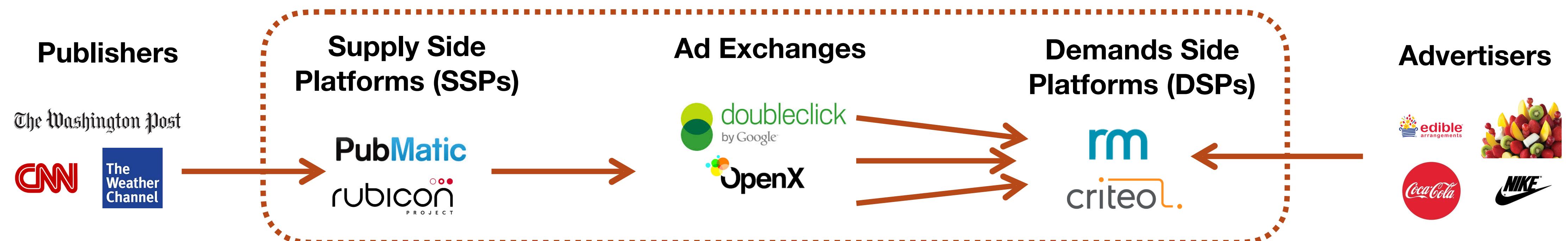
Process of buying and selling ad impressions through real-time auctions.



- SSPs work with publishers to maximize their revenue.

The Rise of Real Time Bidding (RTB)

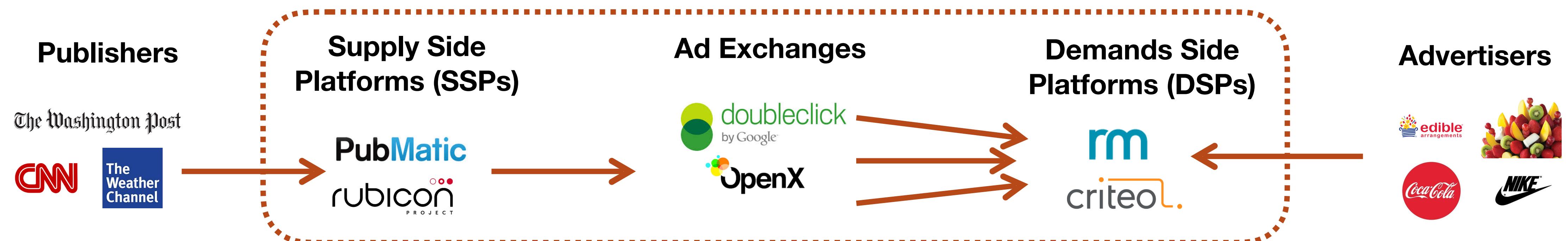
Process of buying and selling ad impressions through real-time auctions.



- SSPs work with publishers to maximize their revenue.
- Ad Exchanges run auctions.

The Rise of Real Time Bidding (RTB)

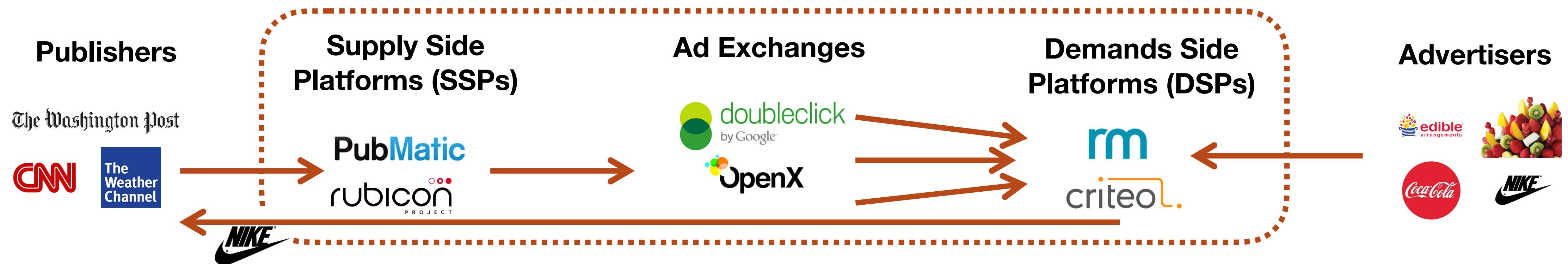
Process of buying and selling ad impressions through real-time auctions.



- SSPs work with publishers to maximize their revenue.
- Ad Exchanges run auctions.
- DSPs place bids on behalf of advertisers.
 - DSPs can analyze context and can bid according to the information they have.

The Rise of Real Time Bidding (RTB)

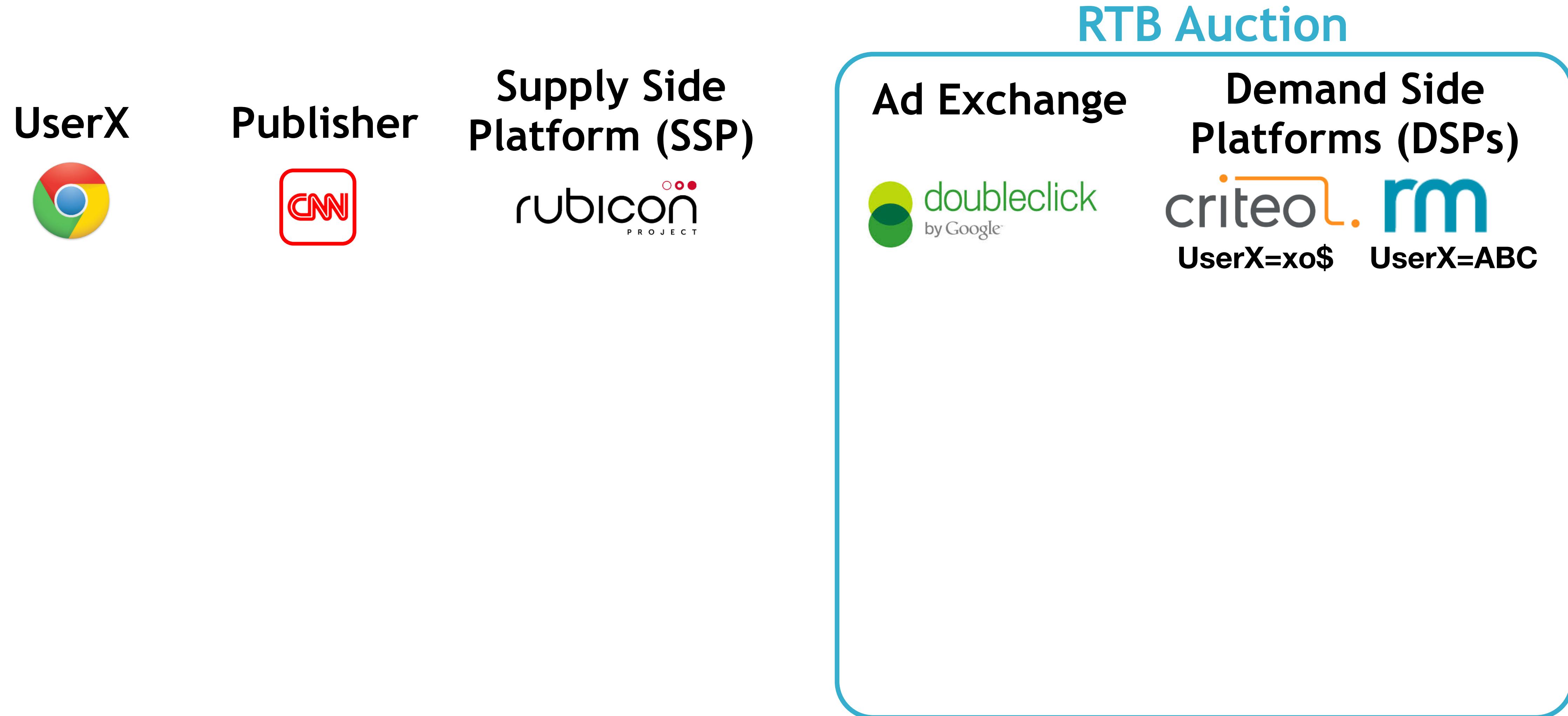
Process of buying and selling ad impressions through real-time auctions.



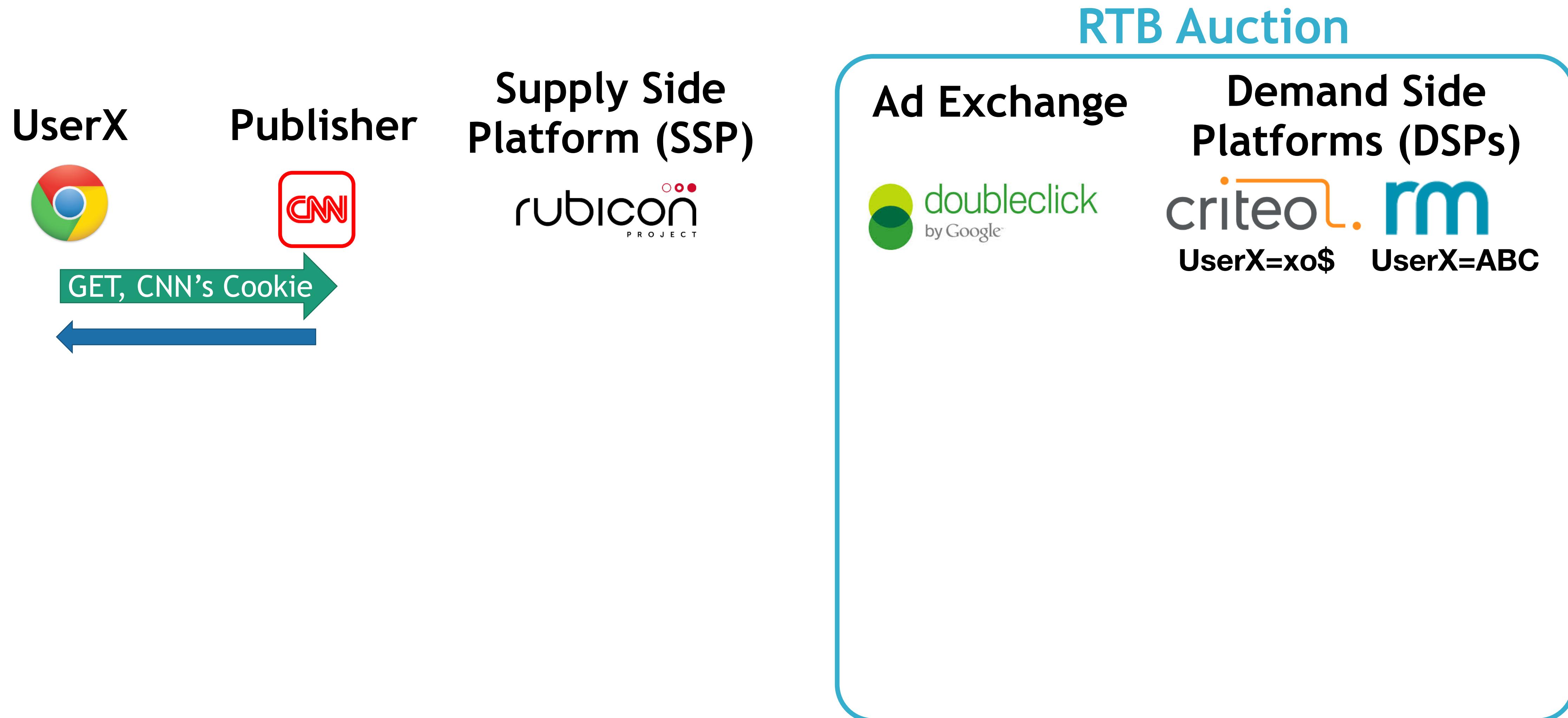
- SSPs work with publishers to maximize their revenue.
- Ad Exchanges run auctions.
- DSPs place bids on behalf of advertisers.
 - DSPs can analyze context and can bid according to the information they have.

1: http://contentb-na2.emarketer.com/content3/US_Programmatic_Ad_Spending_Forecast_2018_eMarketer.pdf

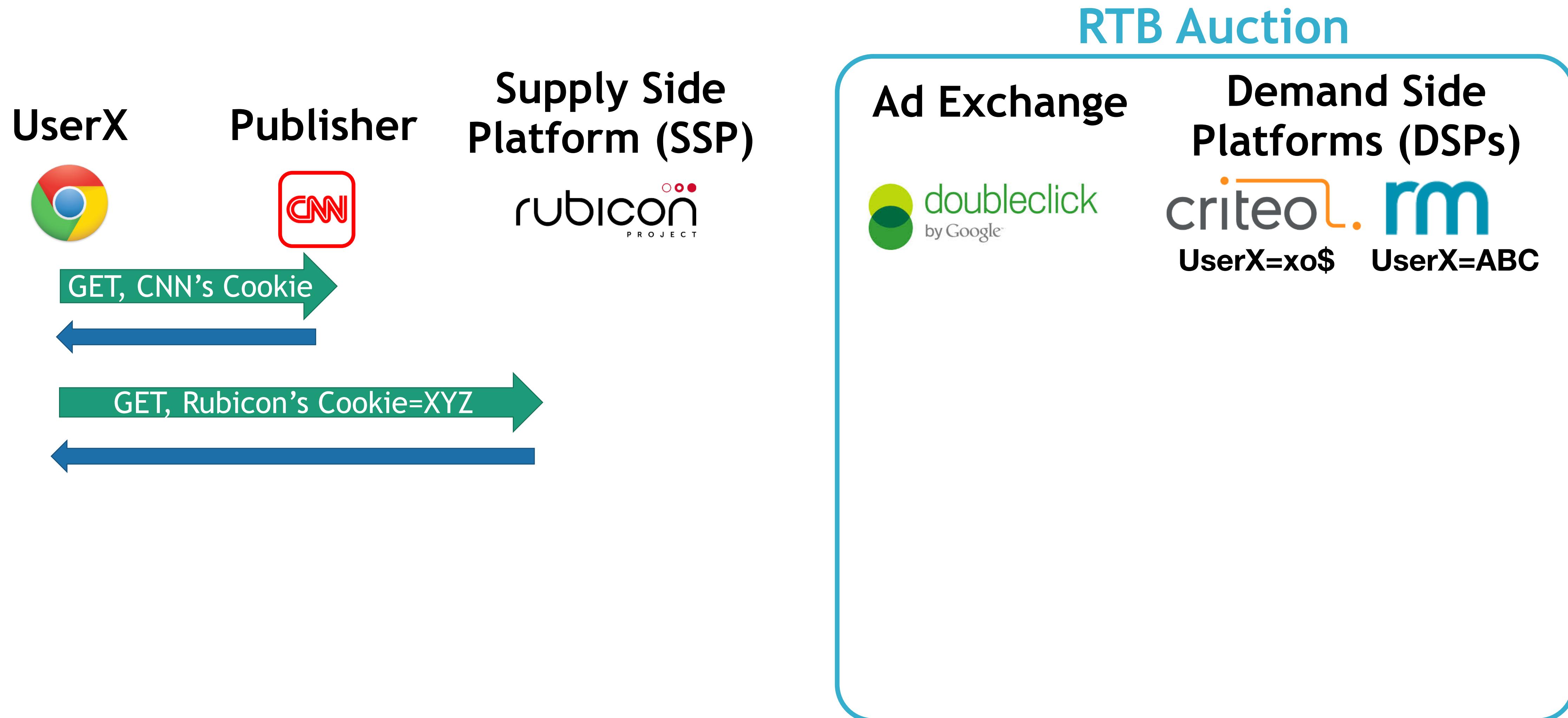
An Example of Real Time Bidding (RTB) Auction



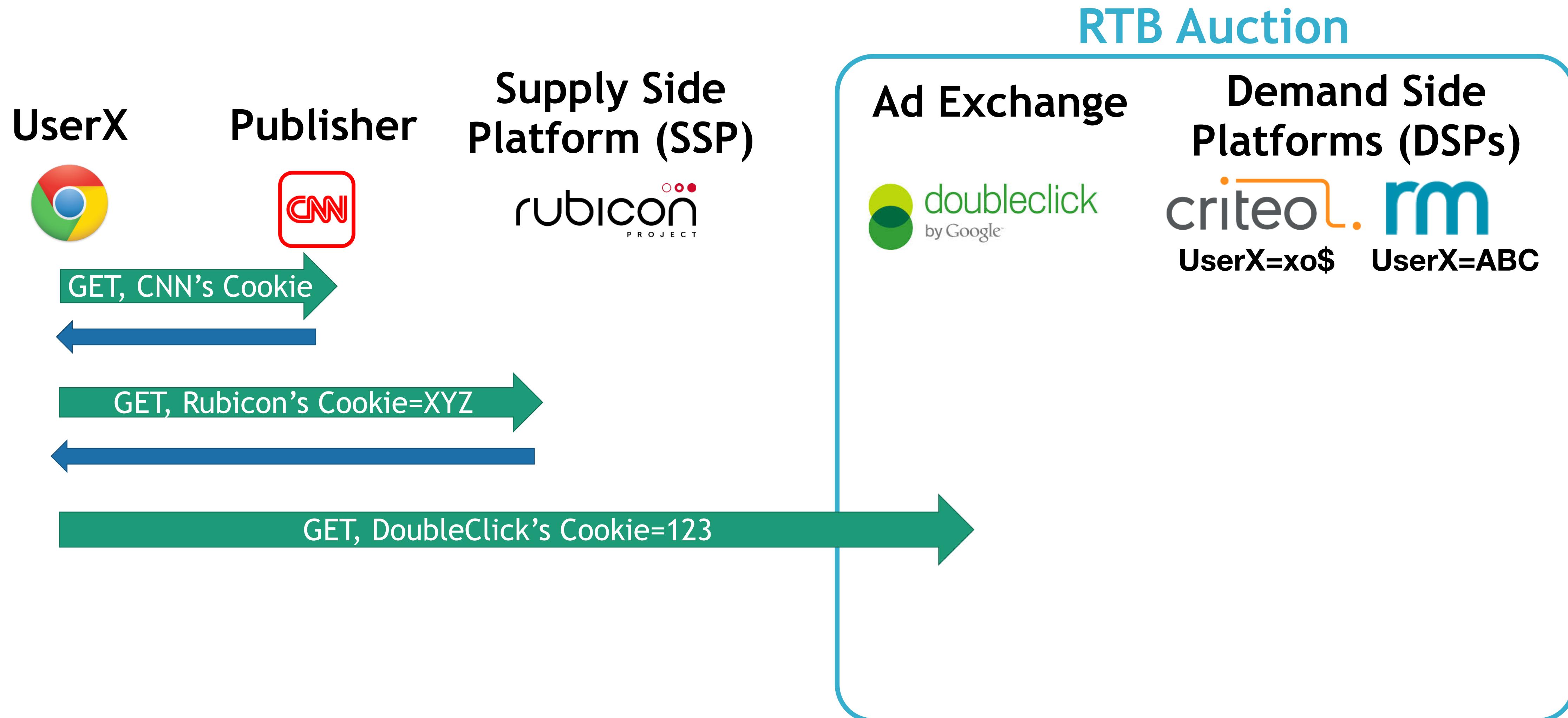
An Example of Real Time Bidding (RTB) Auction



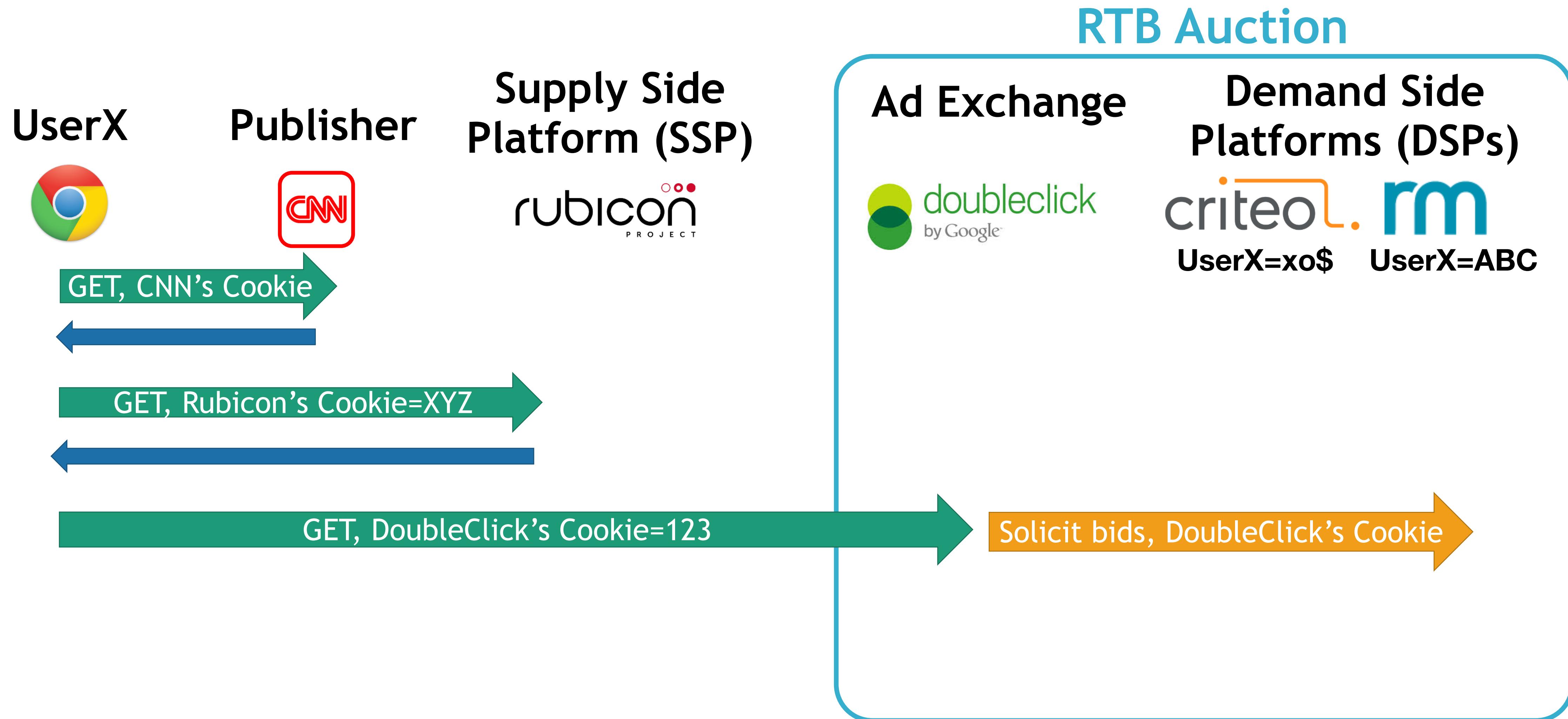
An Example of Real Time Bidding (RTB) Auction



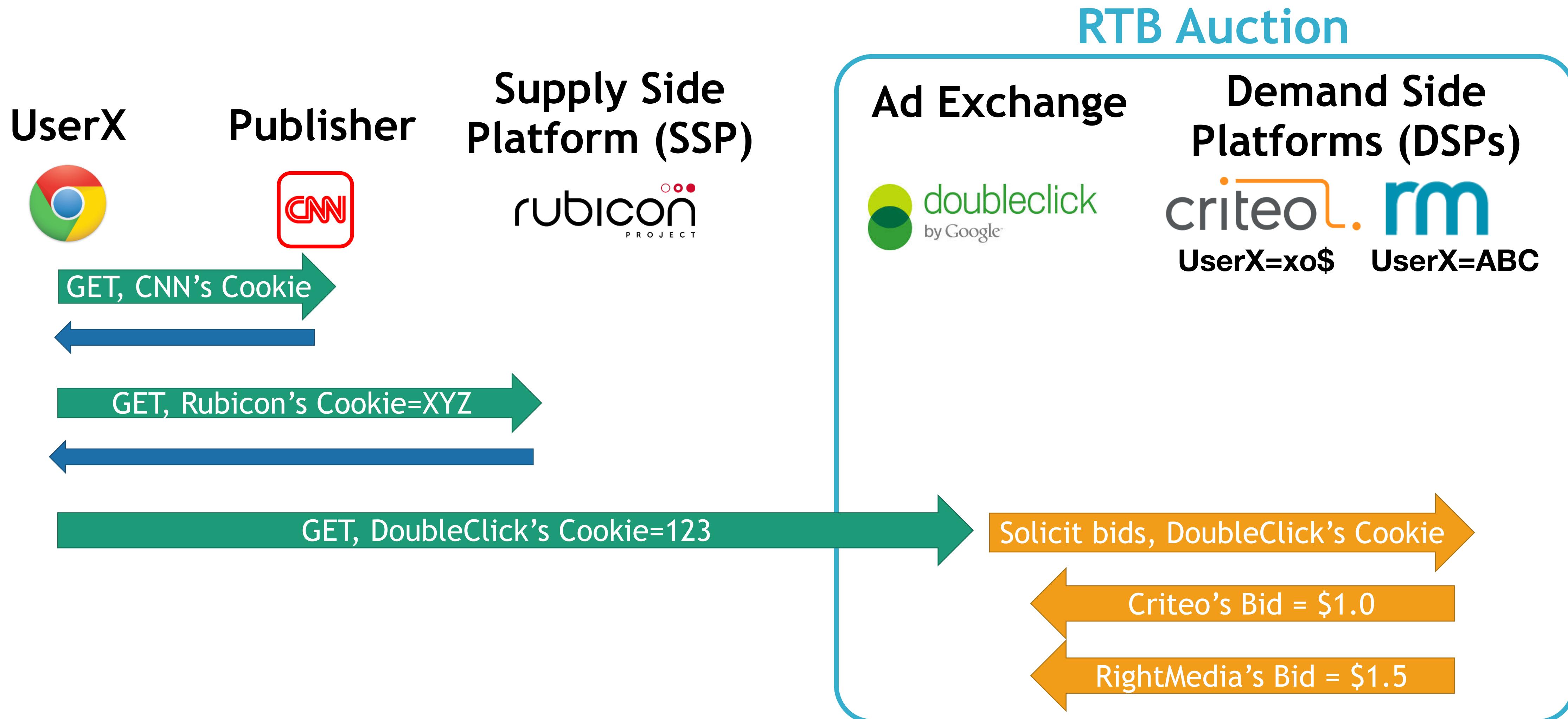
An Example of Real Time Bidding (RTB) Auction



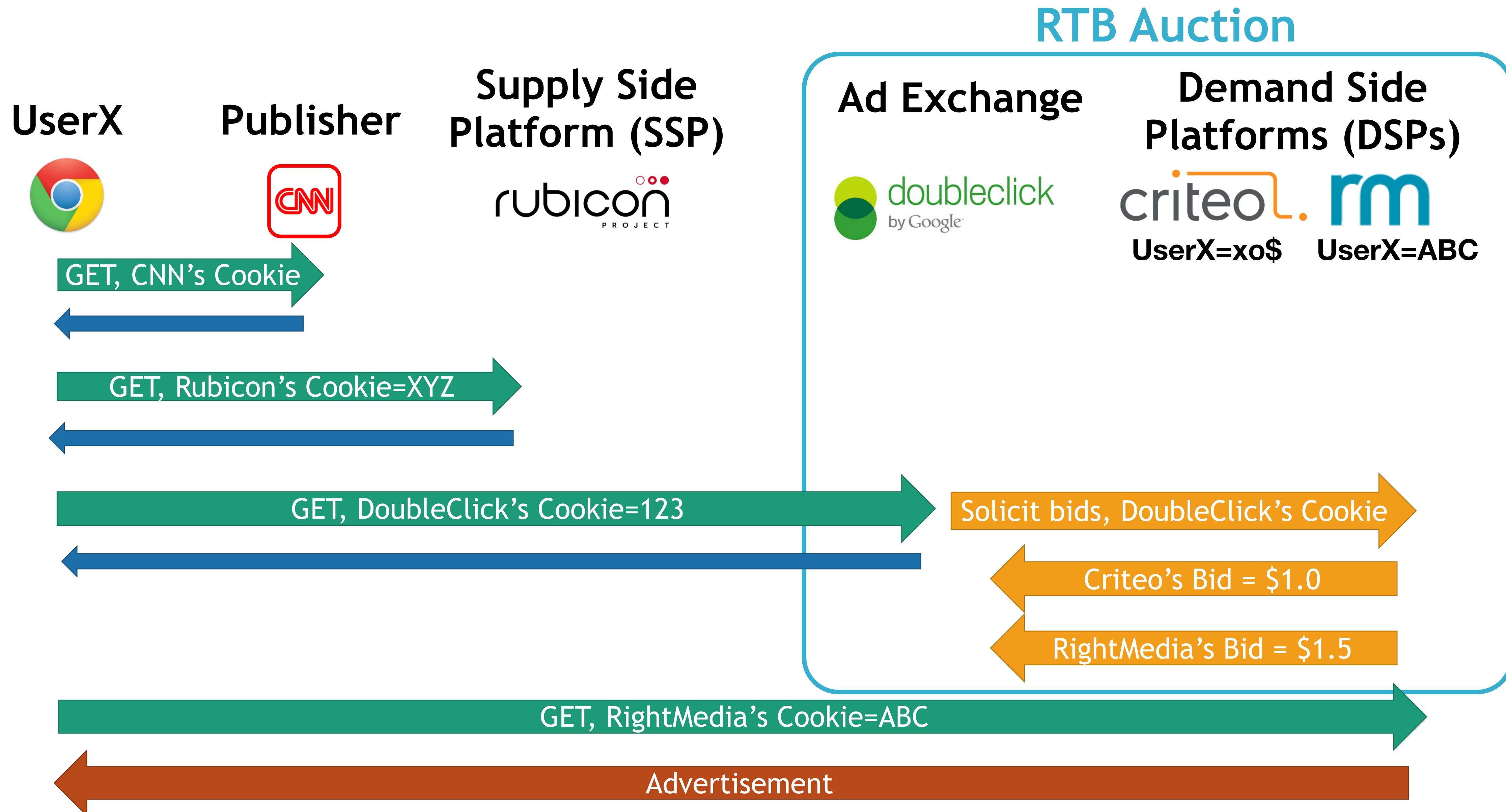
An Example of Real Time Bidding (RTB) Auction



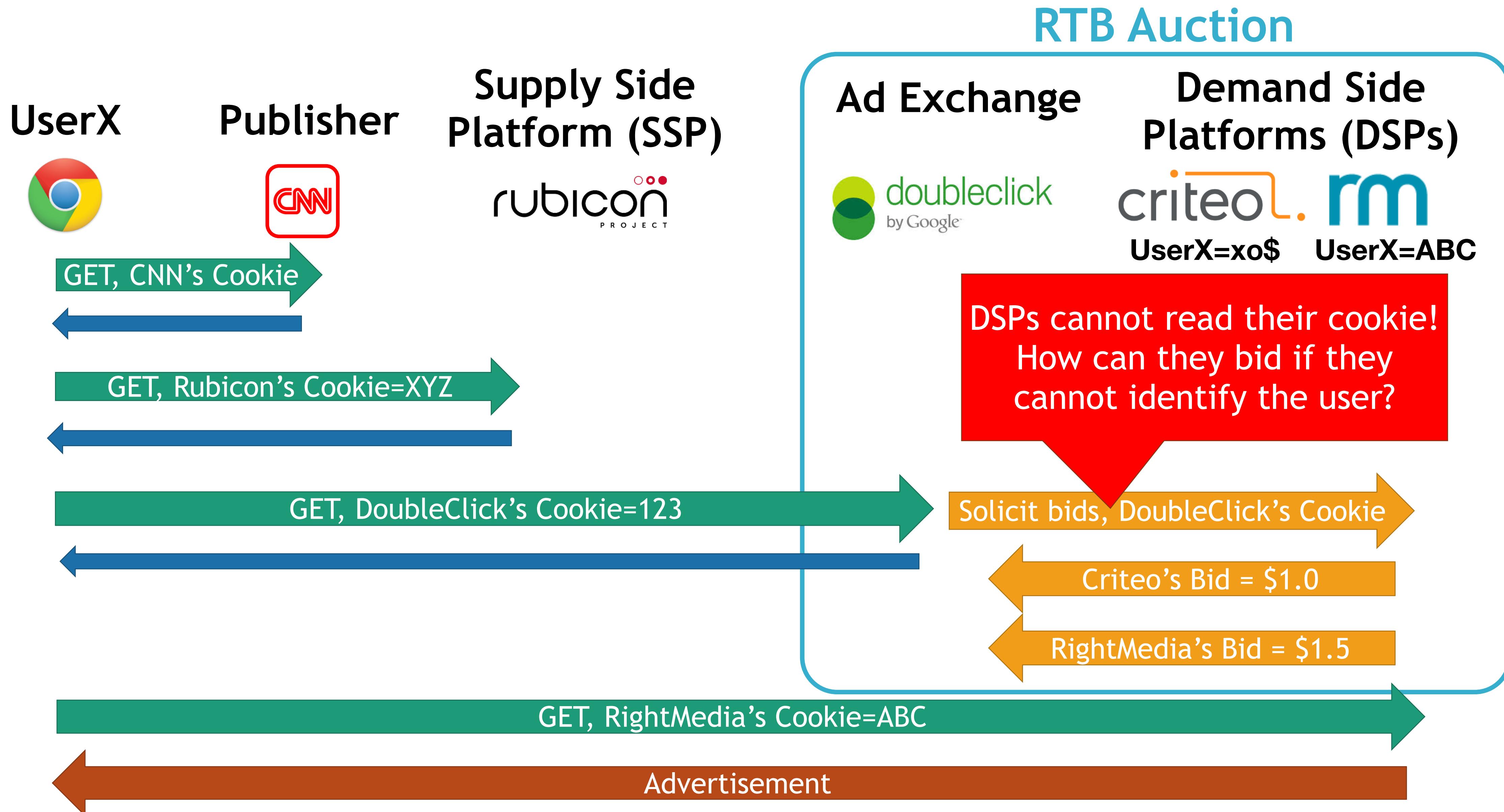
An Example of Real Time Bidding (RTB) Auction



An Example of Real Time Bidding (RTB) Auction



An Example of Real Time Bidding (RTB) Auction



Cookie Matching

Key problem: DSPs cannot read their cookies in the RTB auction

- How can they submit reasonable bids if they cannot identify the user?

Cookie Matching

Key problem: DSPs cannot read their cookies in the RTB auction

- How can they submit reasonable bids if they cannot identify the user?

Solution: [cookie matching](#)

- Also known as cookie synching
- Process of linking the identifiers used by two ad domains

Cookie Matching

Key problem: DSPs cannot read their cookies in the RTB auction

- How can they submit reasonable bids if they cannot identify the user?

Solution: [cookie matching](#)

- Also known as cookie synching
- Process of linking the identifiers used by two ad domains



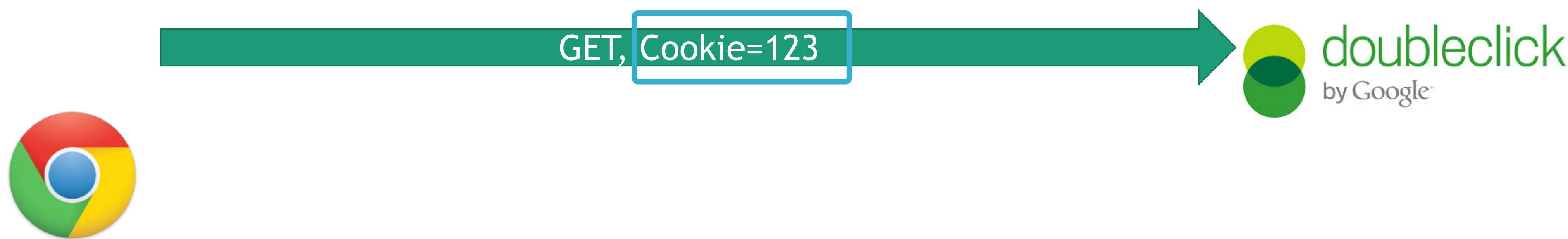
Cookie Matching

Key problem: DSPs cannot read their cookies in the RTB auction

- How can they submit reasonable bids if they cannot identify the user?

Solution: [cookie matching](#)

- Also known as cookie synching
- Process of linking the identifiers used by two ad domains



Cookie Matching

Key problem: DSPs cannot read their cookies in the RTB auction

- How can they submit reasonable bids if they cannot identify the user?

Solution: cookie matching

- Also known as cookie synching
- Process of linking the identifiers used by two ad domains



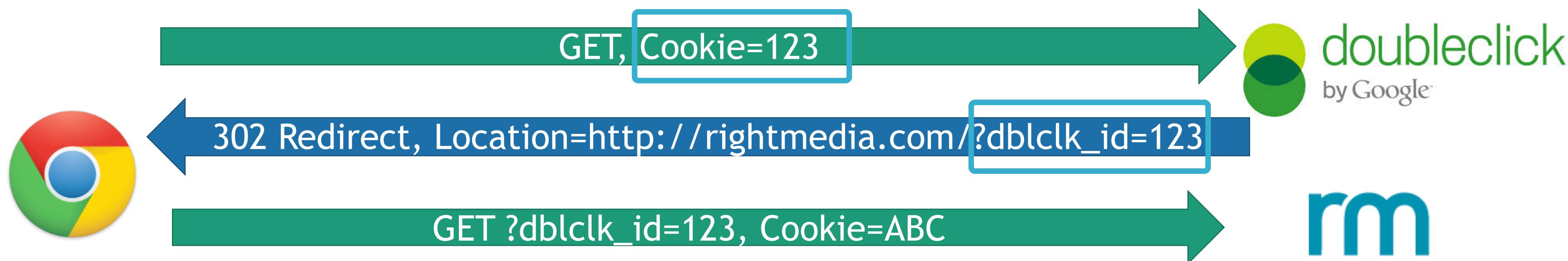
Cookie Matching

Key problem: DSPs cannot read their cookies in the RTB auction

- How can they submit reasonable bids if they cannot identify the user?

Solution: cookie matching

- Also known as cookie synching
- Process of linking the identifiers used by two ad domains



Cookie Matching

Key problem: DSPs cannot read their cookies in the RTB auction

- How can they submit reasonable bids if they cannot identify the user?

Solution: cookie matching

- Also known as cookie synching
- Process of linking the identifiers used by two ad domains



Cookie Matching

Key problem: DSPs cannot read their cookies in the RTB auction

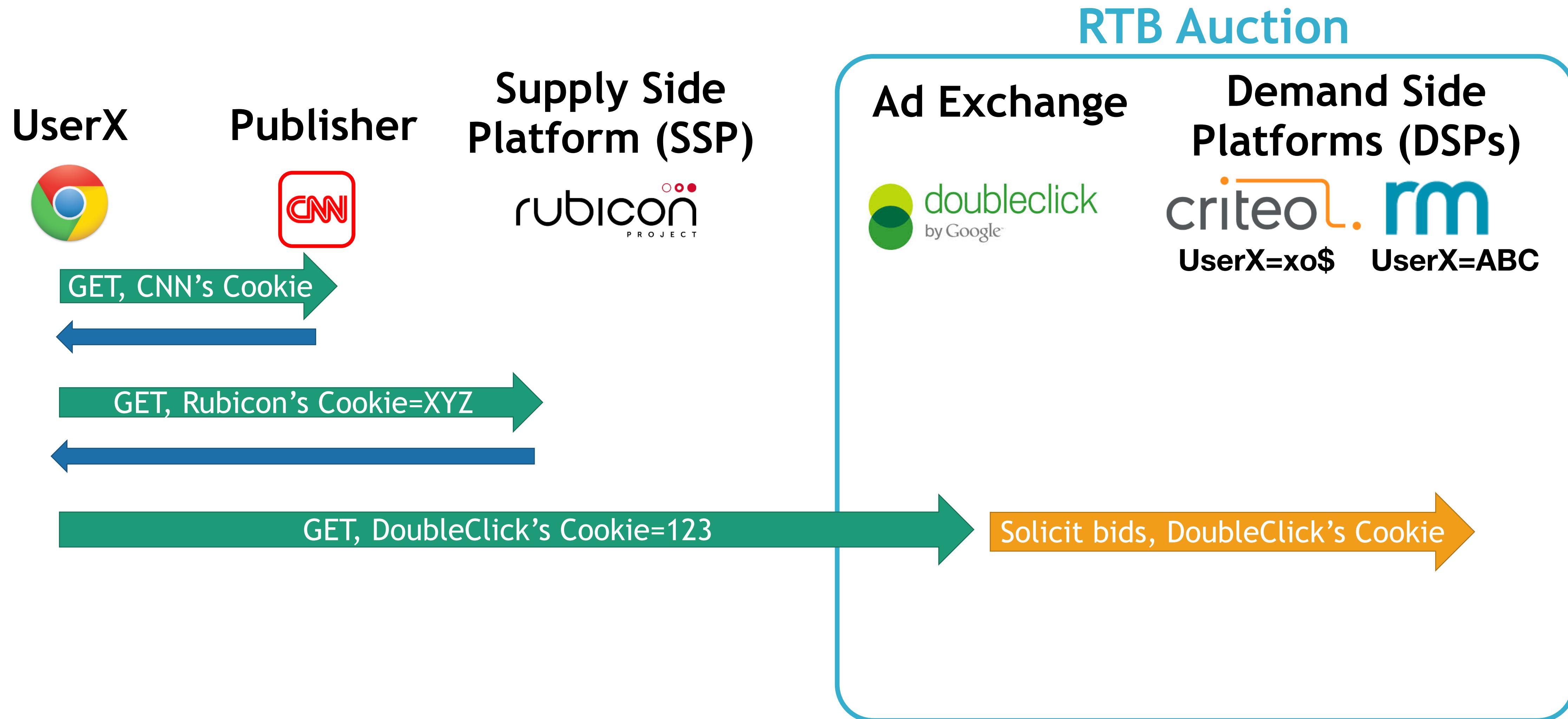
- How can they submit reasonable bids if they cannot identify the user?

Solution: cookie matching

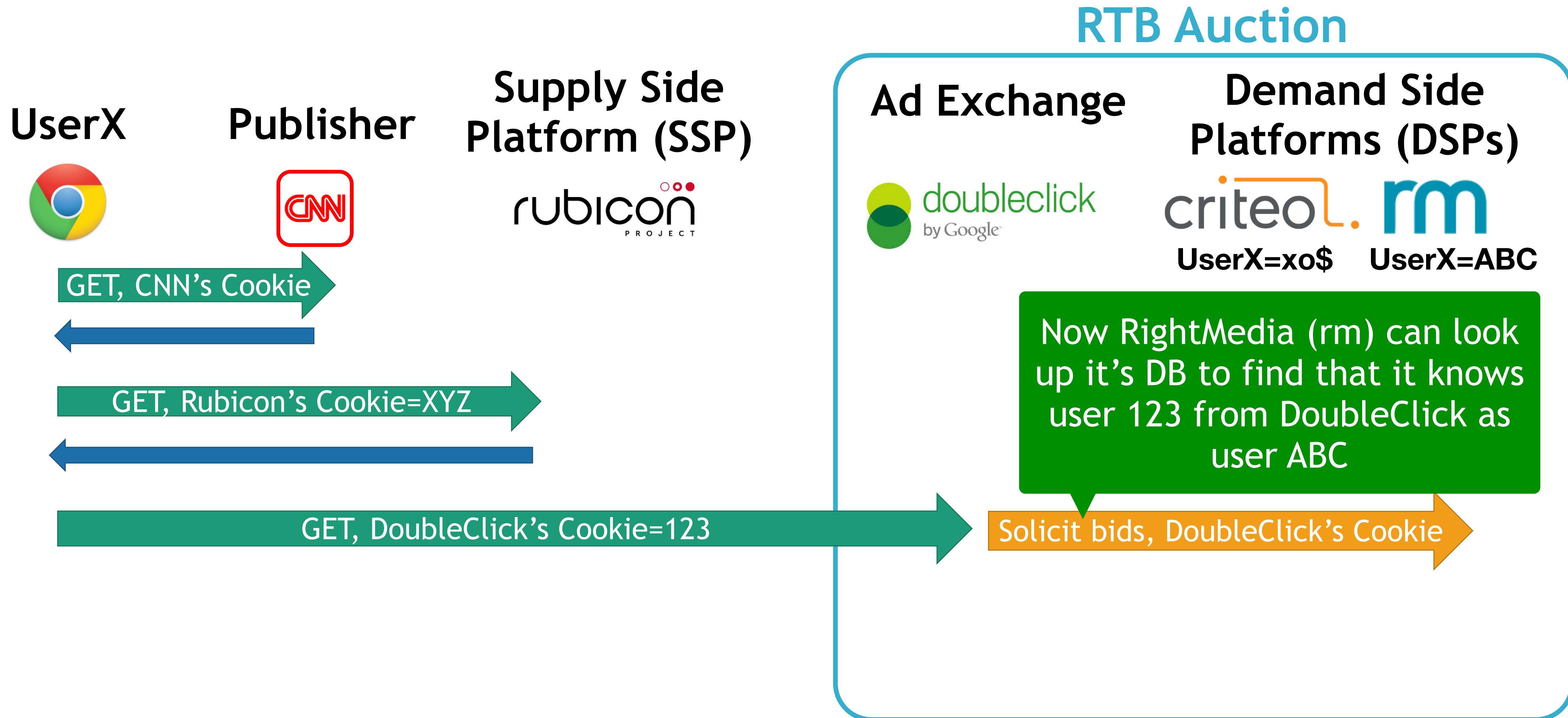
- Also known as cookie synching
- Process of linking the identifiers used by two ad domains



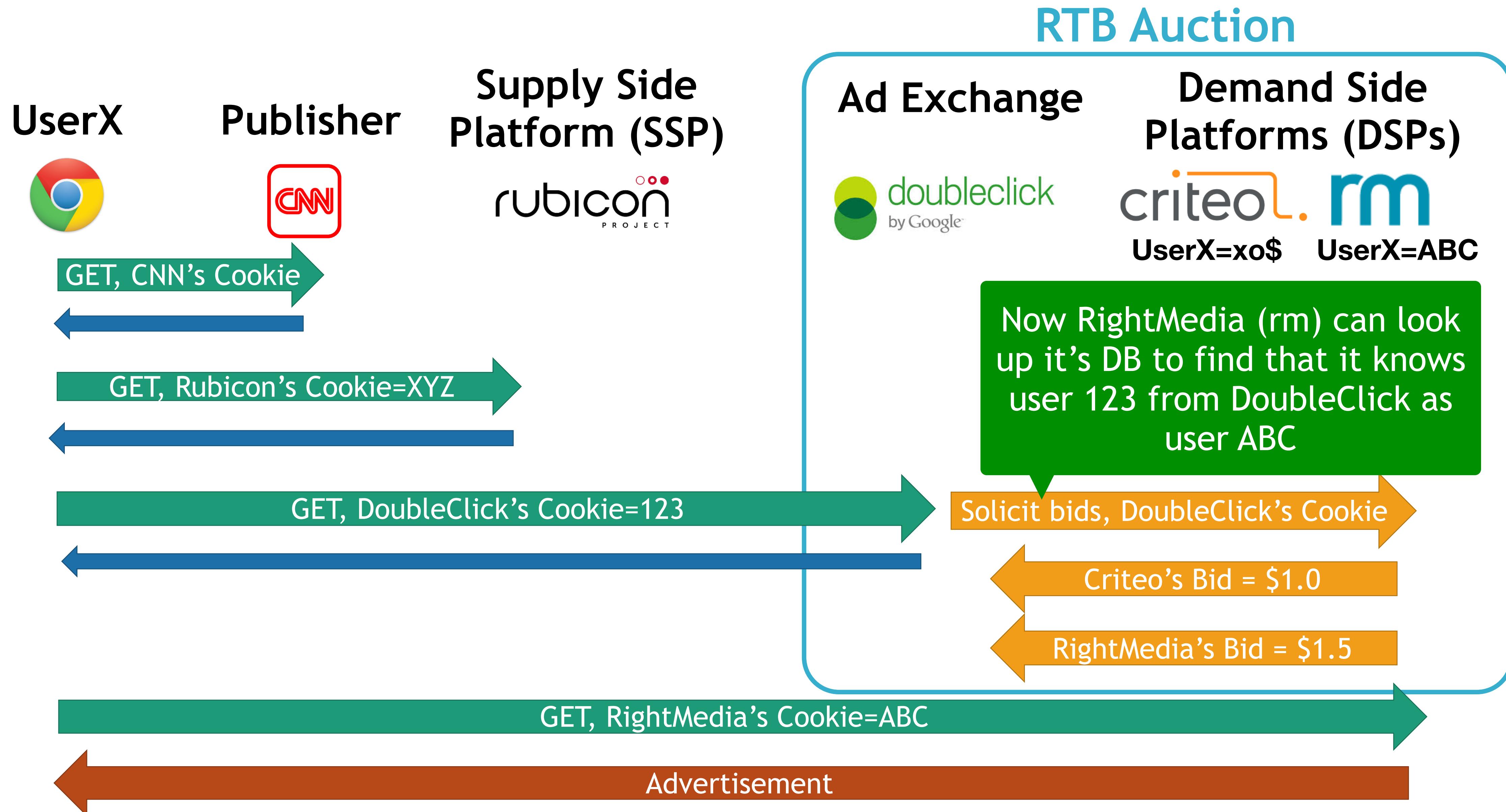
An Example of Real Time Bidding (RTB) Auction



An Example of Real Time Bidding (RTB) Auction

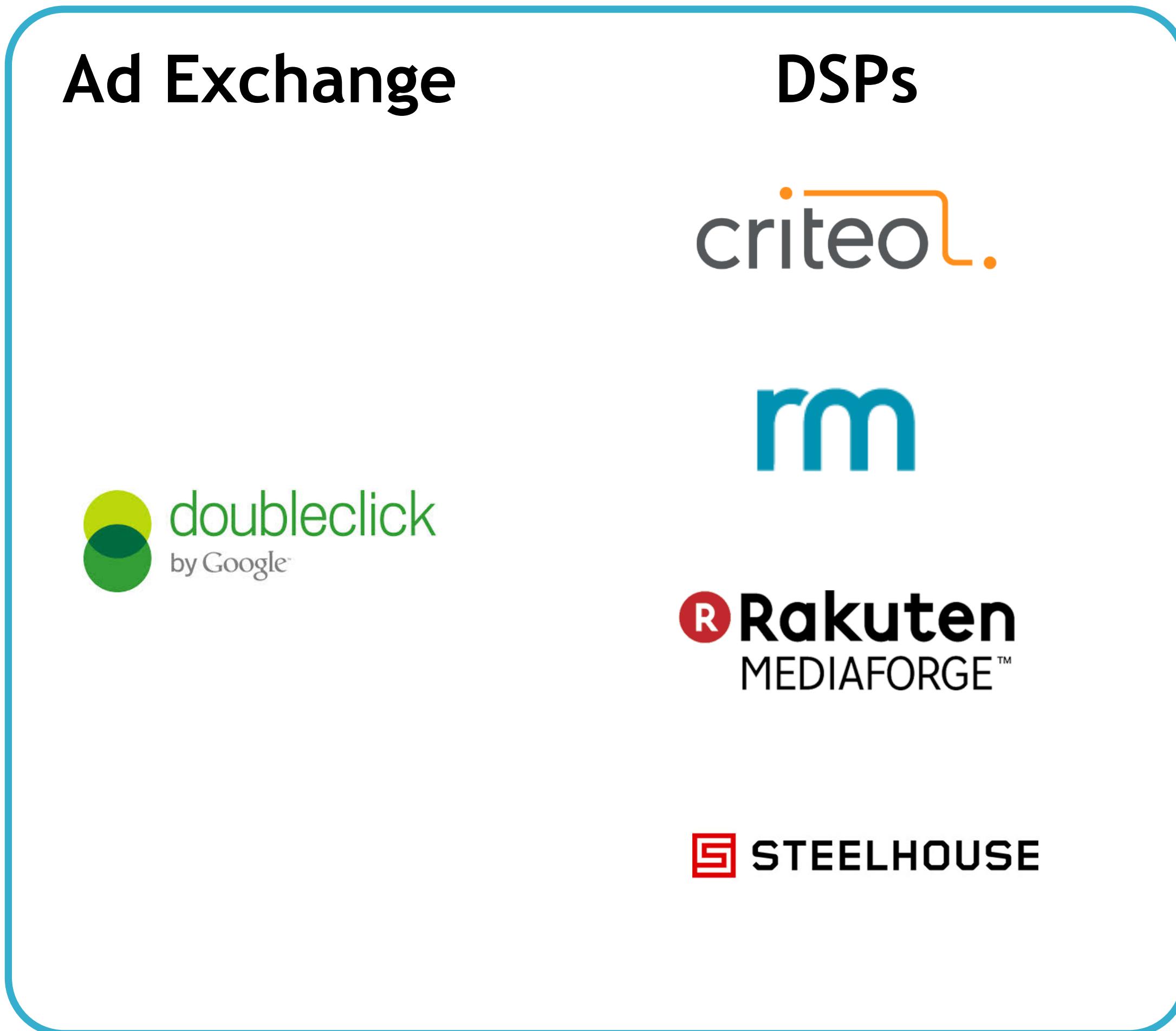


An Example of Real Time Bidding (RTB) Auction



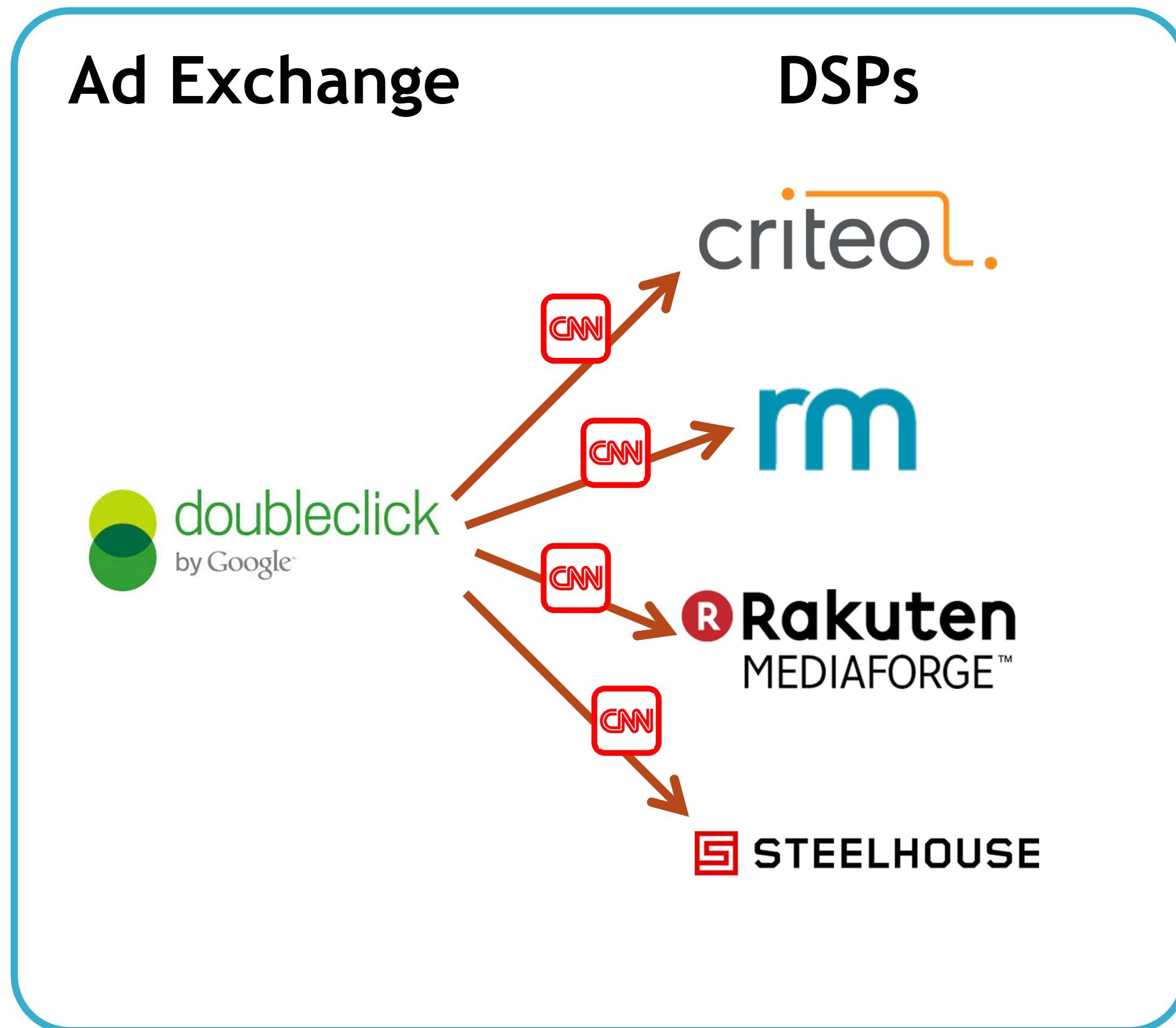
Close Collaboration Among Ad Networks

RTB Auction



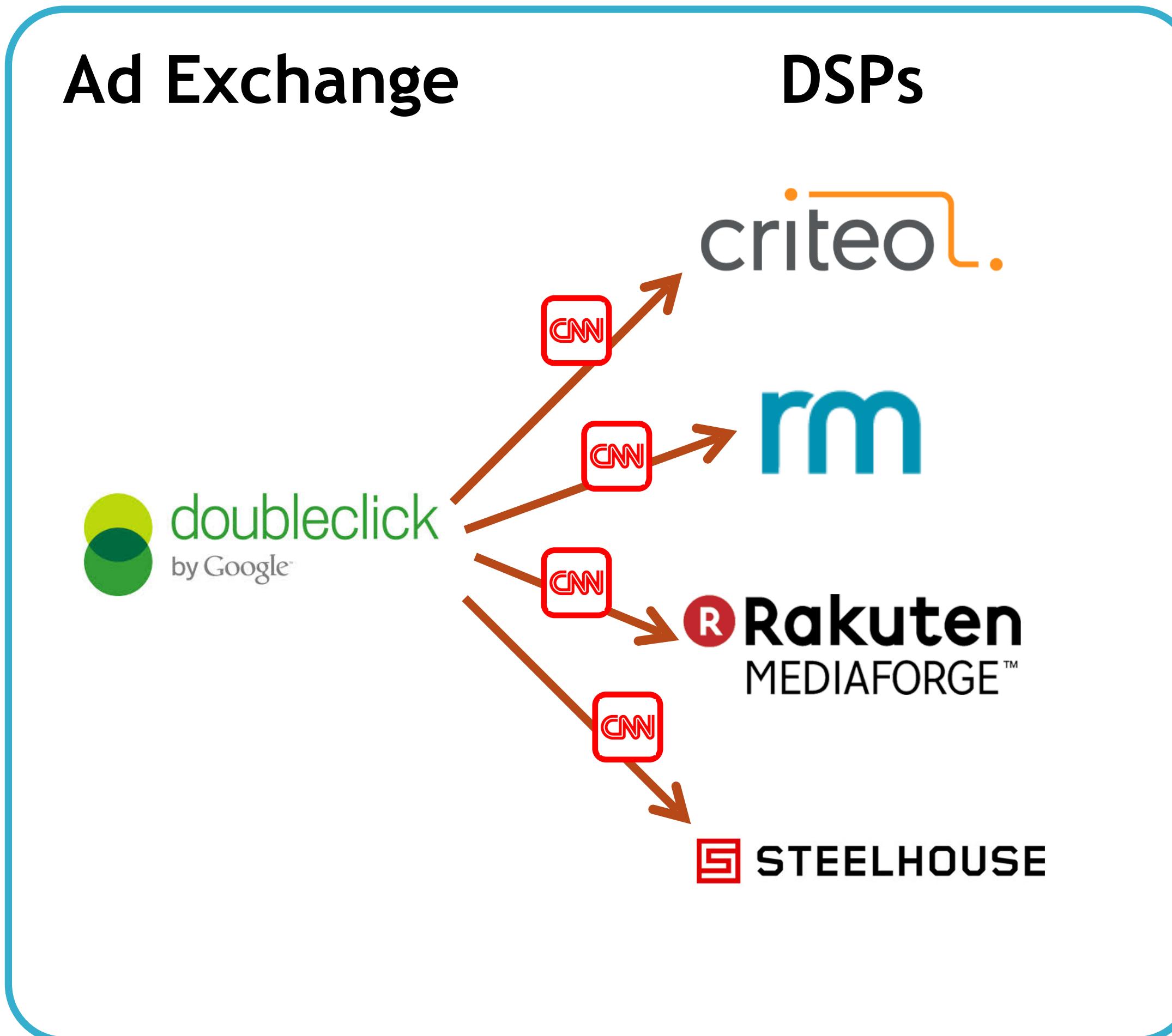
Close Collaboration Among Ad Networks

RTB Auction



Close Collaboration Among Ad Networks

RTB Auction

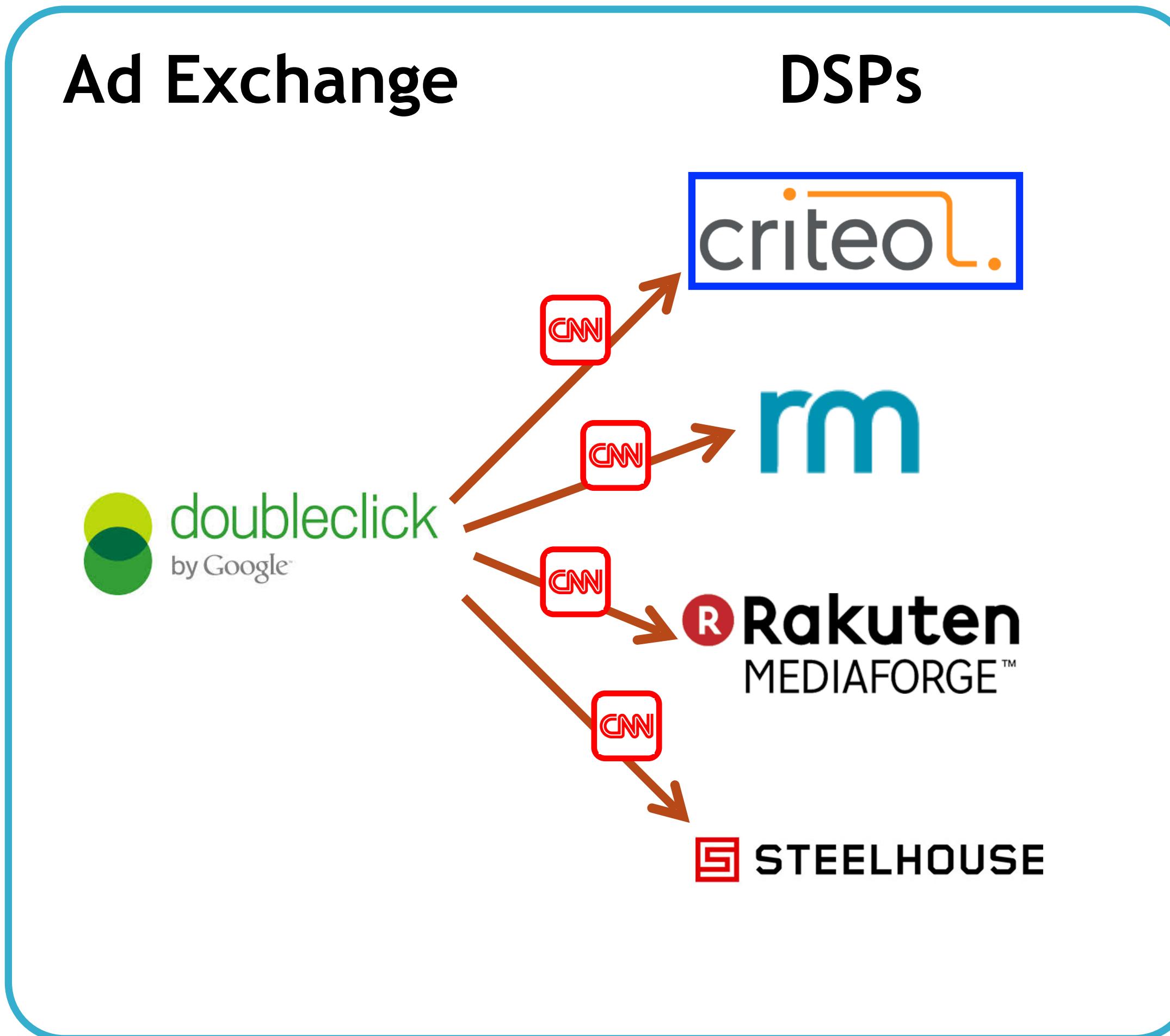


Takeaway #1

- Cookie Matching is a pre-requisite for RTB
- Leads to close collaboration among ad networks

Close Collaboration Among Ad Networks

RTB Auction

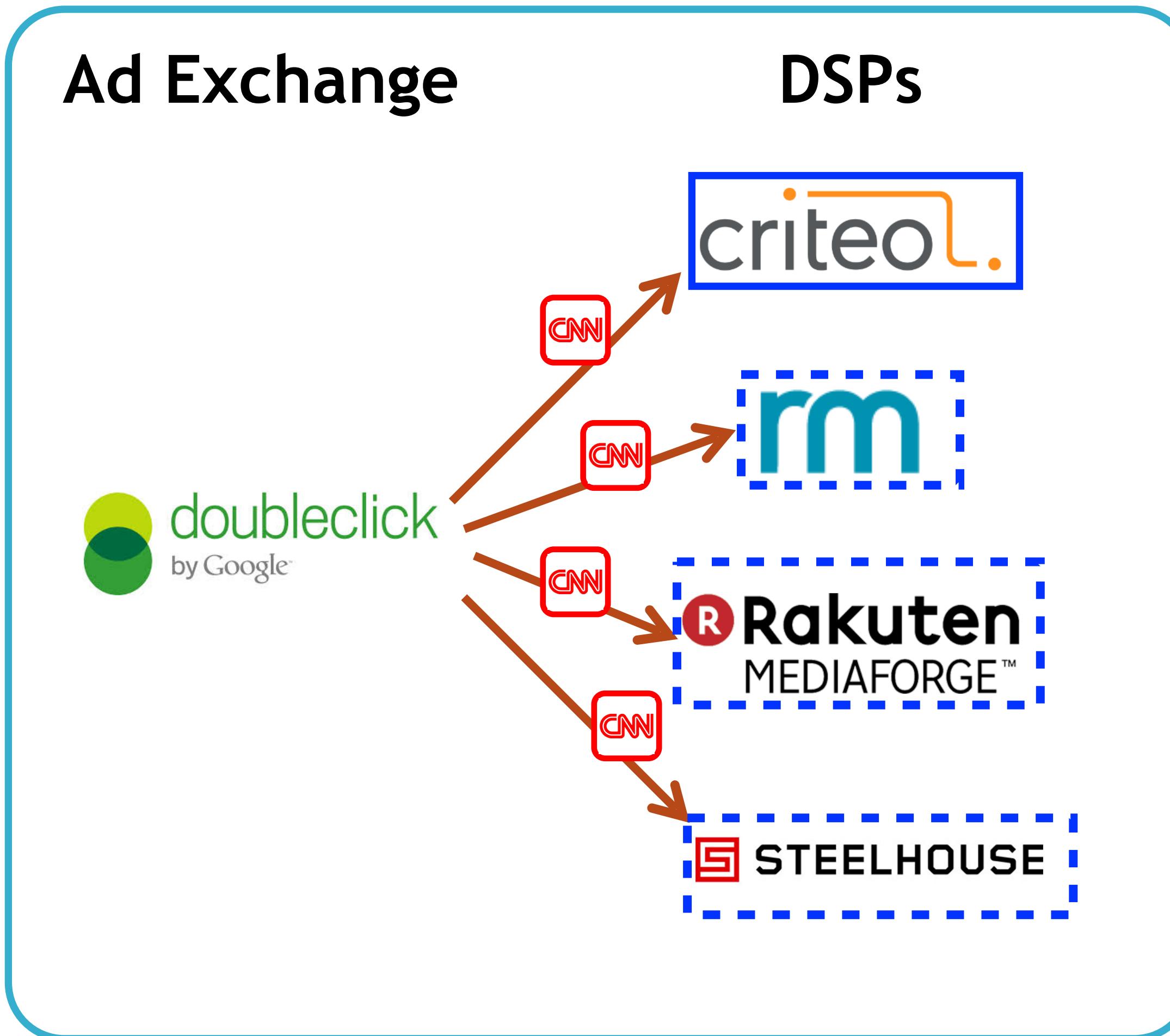


Takeaway #1

- Cookie Matching is a pre-requisite for RTB
- Leads to close collaboration among ad networks

Close Collaboration Among Ad Networks

RTB Auction

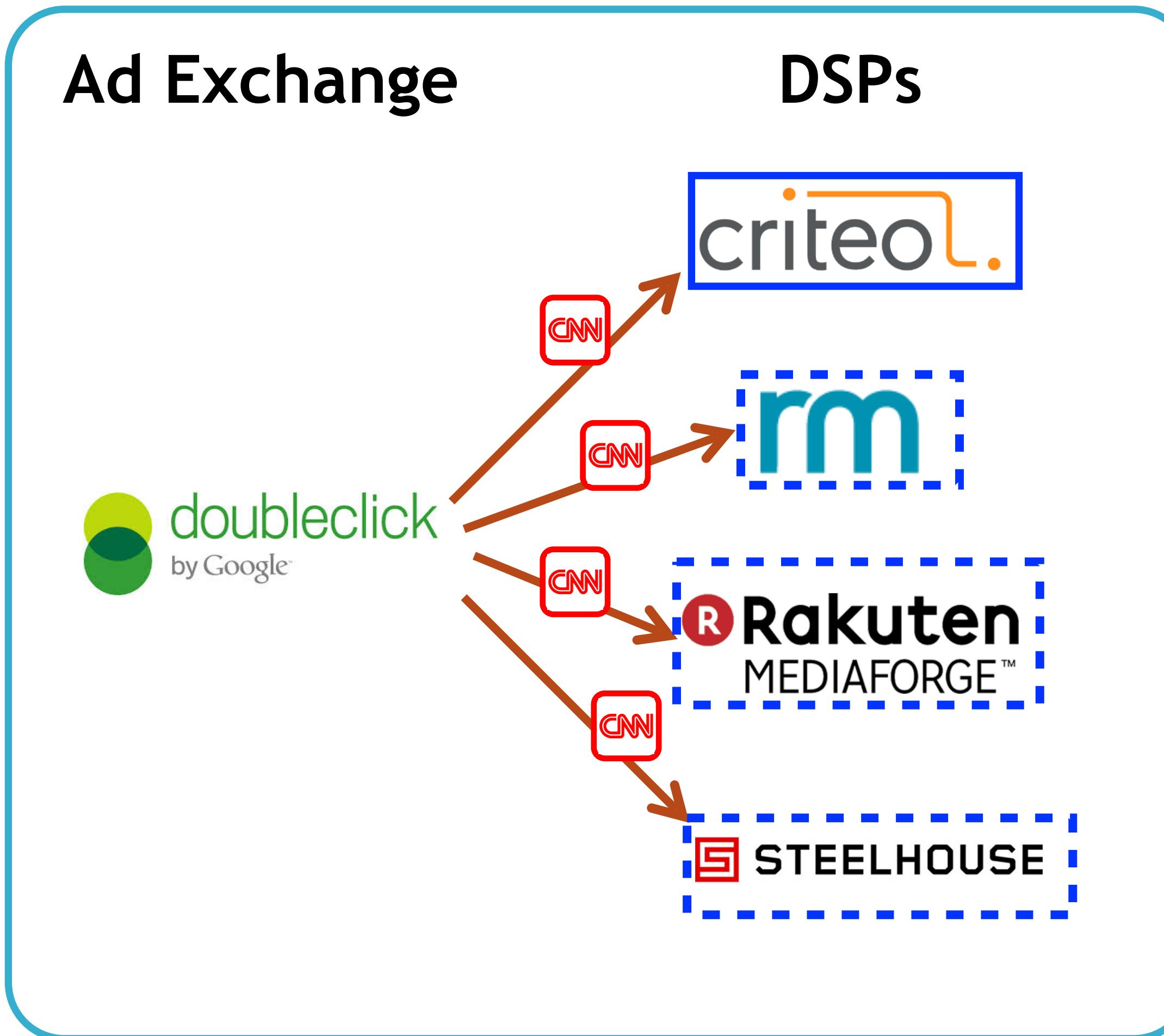


Takeaway #1

- Cookie Matching is a pre-requisite for RTB
- Leads to close collaboration among ad networks

Close Collaboration Among Ad Networks

RTB Auction



Takeaway #1

- Cookie Matching is a pre-requisite for RTB
- Leads to close collaboration among ad networks

Takeaway #2

- In RTB, All DSP partners view user impression
- Losing DSPs also learns some information

Thesis Statement

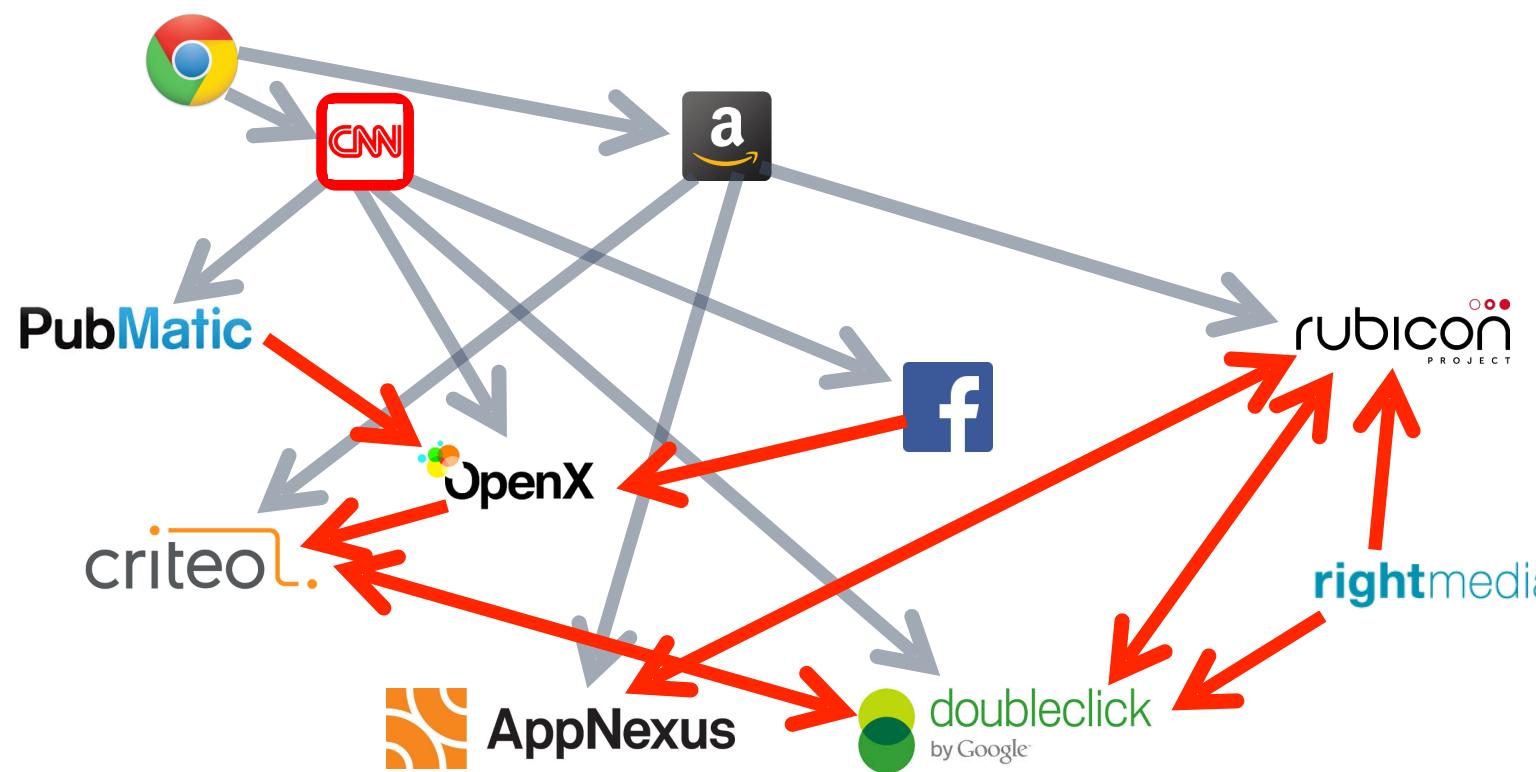
Real Time Bidding has fueled information sharing behind the scenes

- My work focuses on understanding its privacy implications by capturing information sharing

Thesis Statement

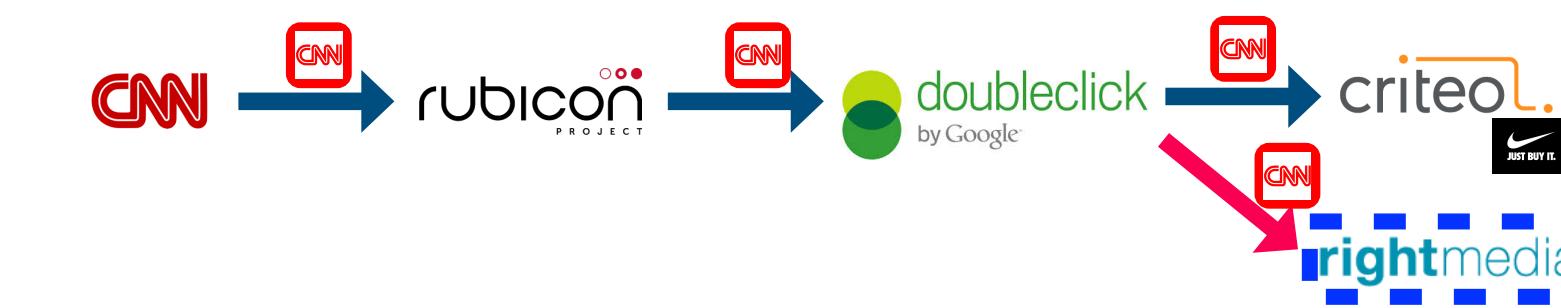
Real Time Bidding has fueled information sharing behind the scenes

- My work focuses on understanding its privacy implications by capturing information sharing



Tracing Information Flows Between Ad Exchanges Using Retargeted Ads

USENIX Security '16



Diffusion of User Tracking Data in the Online Advertising Ecosystem

PETS '18

Outline

Background & Motivation

Limitations of Prior Work

Detecting Information Sharing

- Using Retargeted Ads

Propagation of Tracking Data

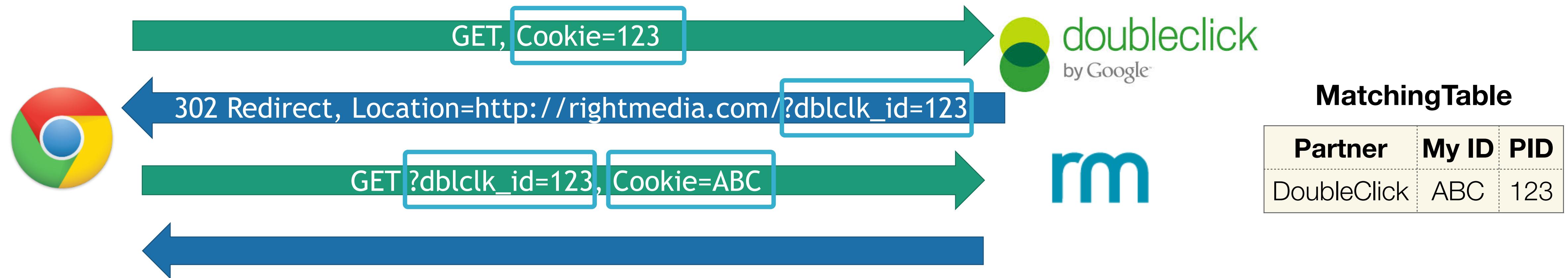
- Modeling via Graph

Ongoing Work

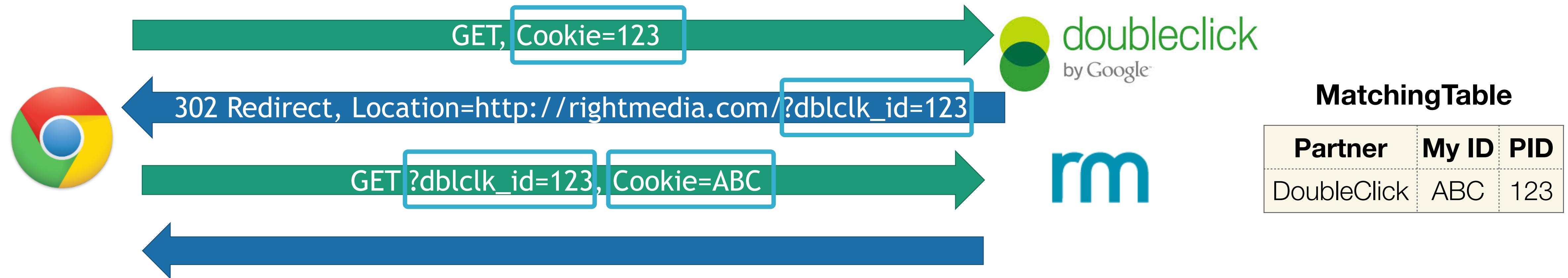
- Cross-device Tracking
- Automation of Adblocking Filter Lists

Research Plan

Using Redirects to Determine Information Sharing?



Using Redirects to Determine Information Sharing?

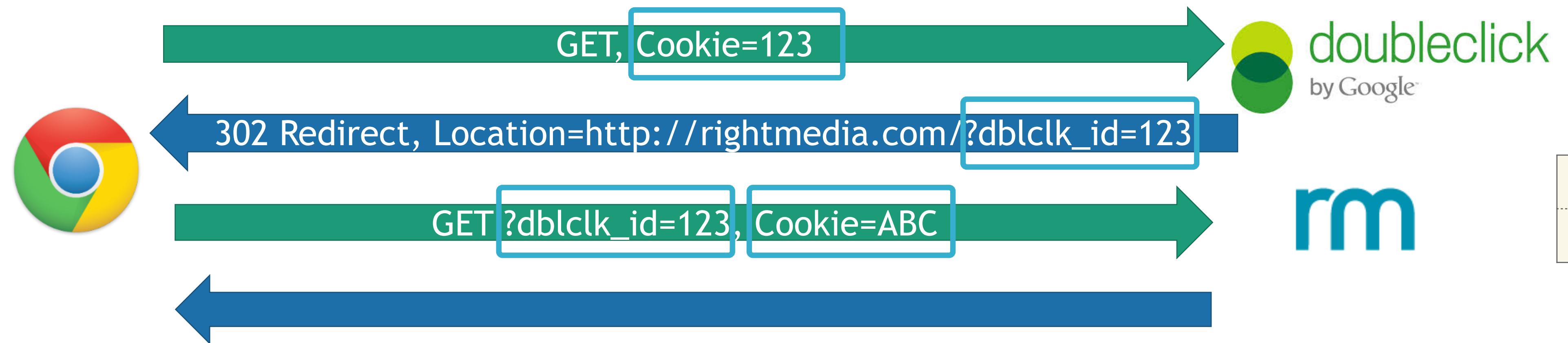


Can't we just rely on HTTP headers to look for redirects?

- Acar *et al.* found hundreds of domains passing identifiers to each other
- Olejnik *et al.* found 125 exchanges matching cookies
- Falahrastegar *et al.* analyzed clusters of exchanges that share the exact same cookies

It's a decent approach — but not without limitations

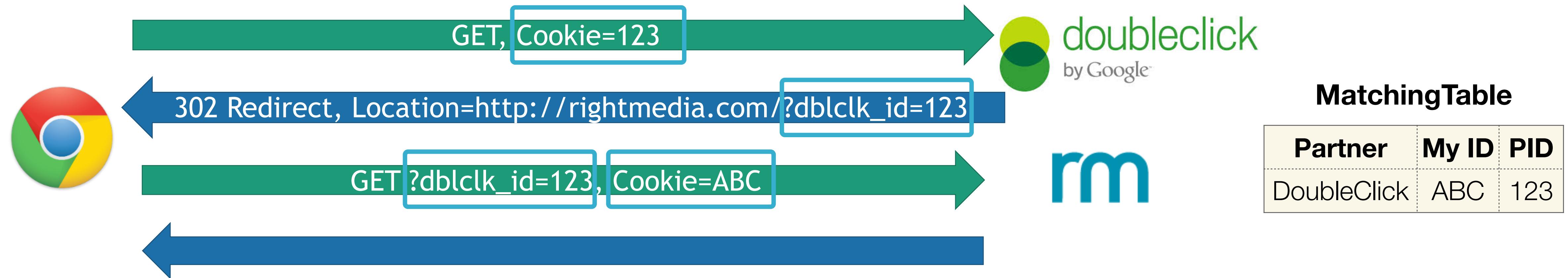
Using Redirects to Determine Information Sharing?



MatchingTable

Partner	My ID	PID
DoubleClick	ABC	123

Using Redirects to Determine Information Sharing?



Can't we just rely on HTTP headers to look for redirects?

- Acar *et al.* found hundreds of domains passing identifiers to each other
- Olejnik *et al.* found 125 exchanges matching cookies
- Falahrastegar *et al.* analyzed clusters of exchanges that share the exact same cookies

It's a decent approach — but not without limitations

Limitation #1: Server Side Matching

Limitation #1: Server Side Matching

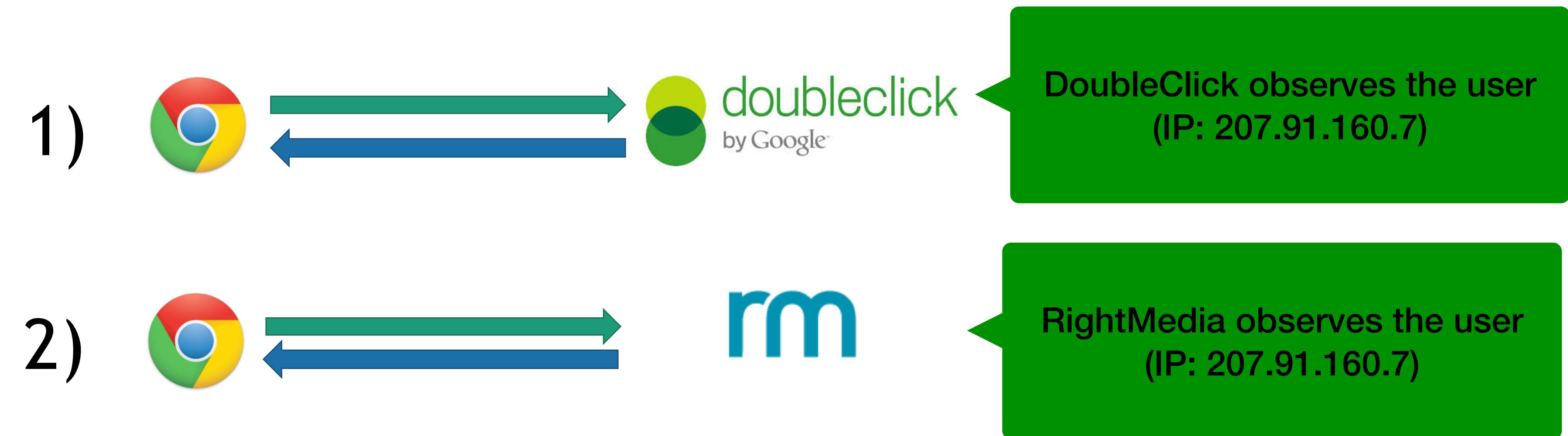
1)



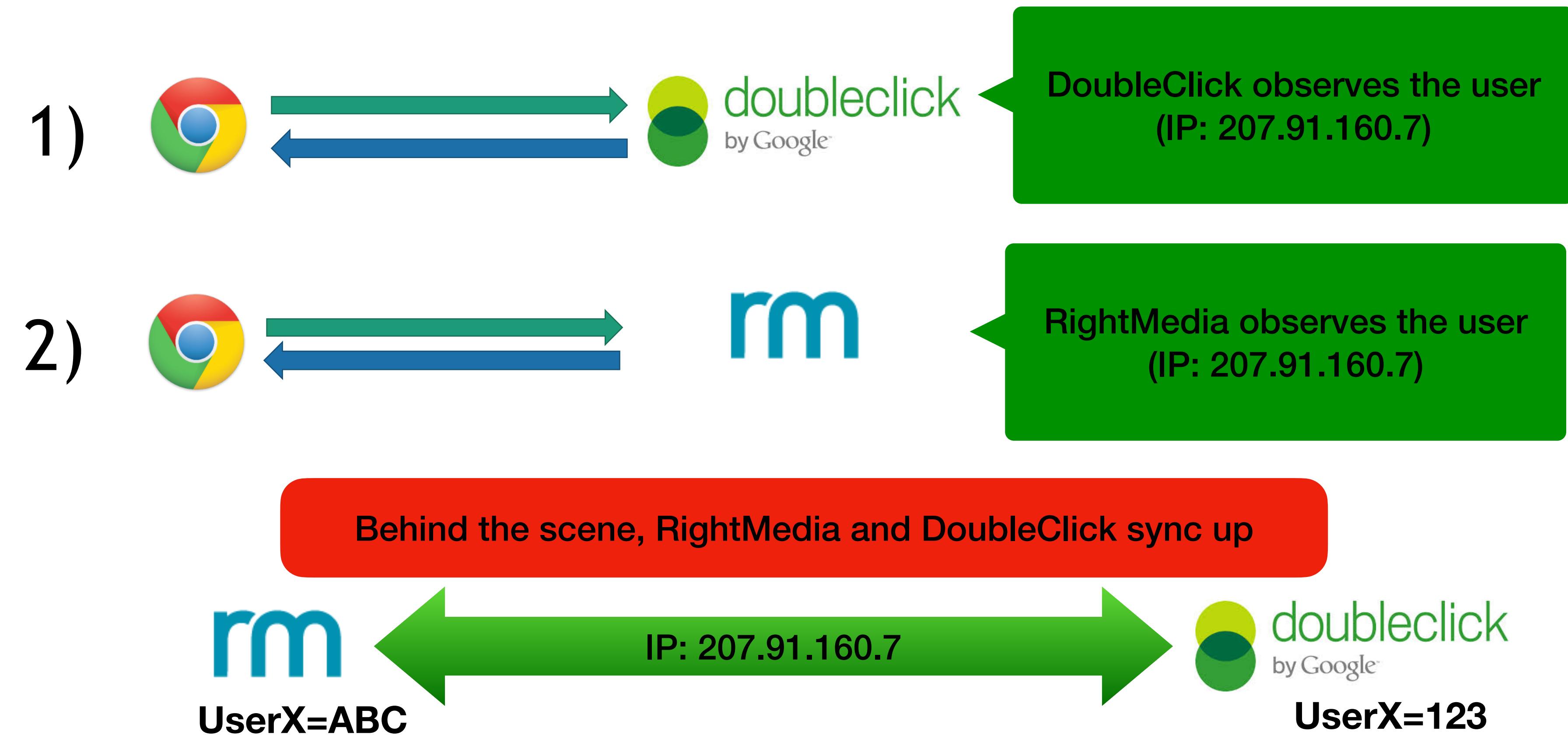
Limitation #1: Server Side Matching



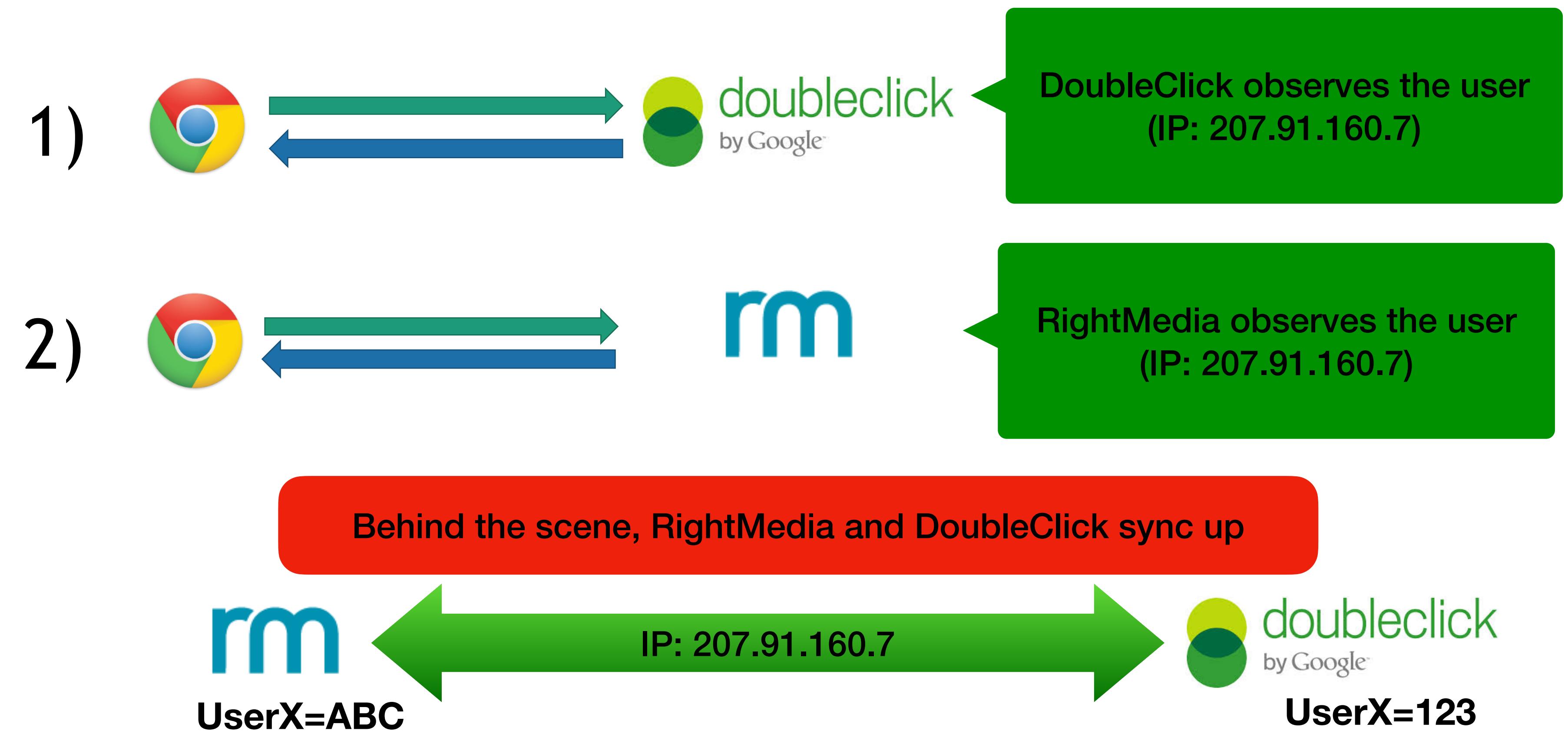
Limitation #1: Server Side Matching



Limitation #1: Server Side Matching



Limitation #1: Server Side Matching



Matching Table

Partner	My ID	PID
DoubleClick	ABC	123

Note: Matching does not happen via the browser (client side)

Limitation #2: Obfuscation

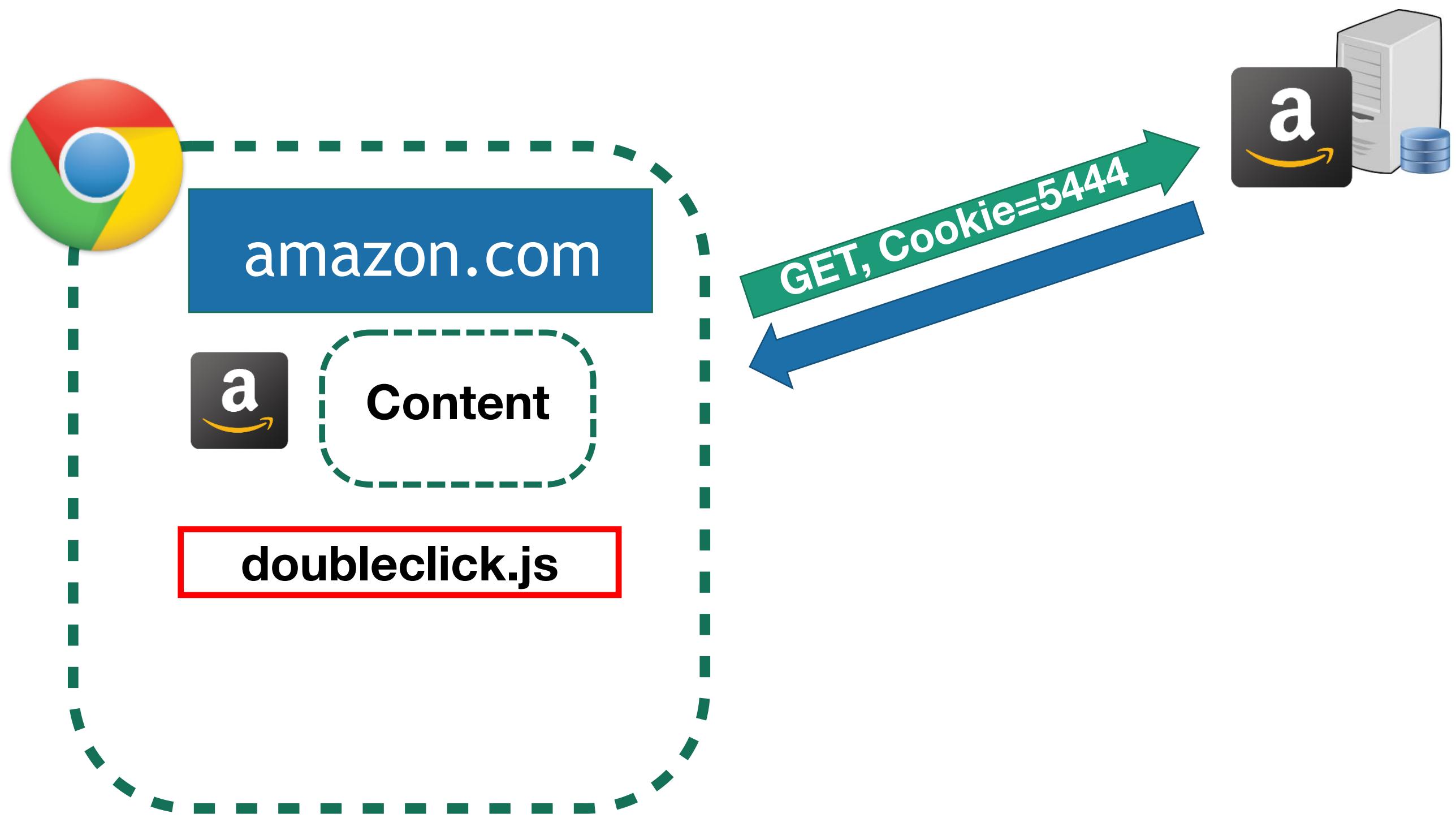
Limitation #2: Obfuscation



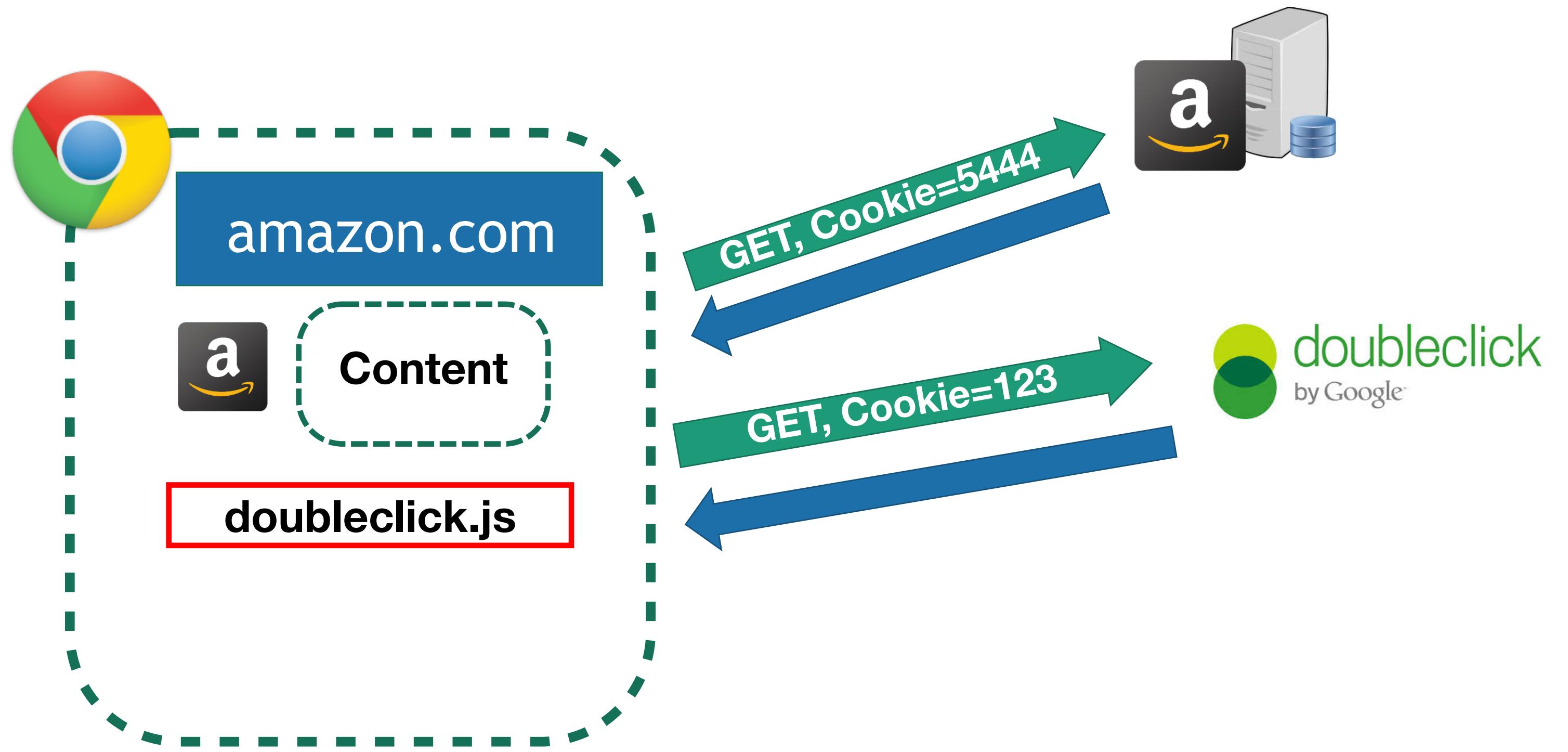
Limitation #2: Obfuscation



Limitation #2: Obfuscation



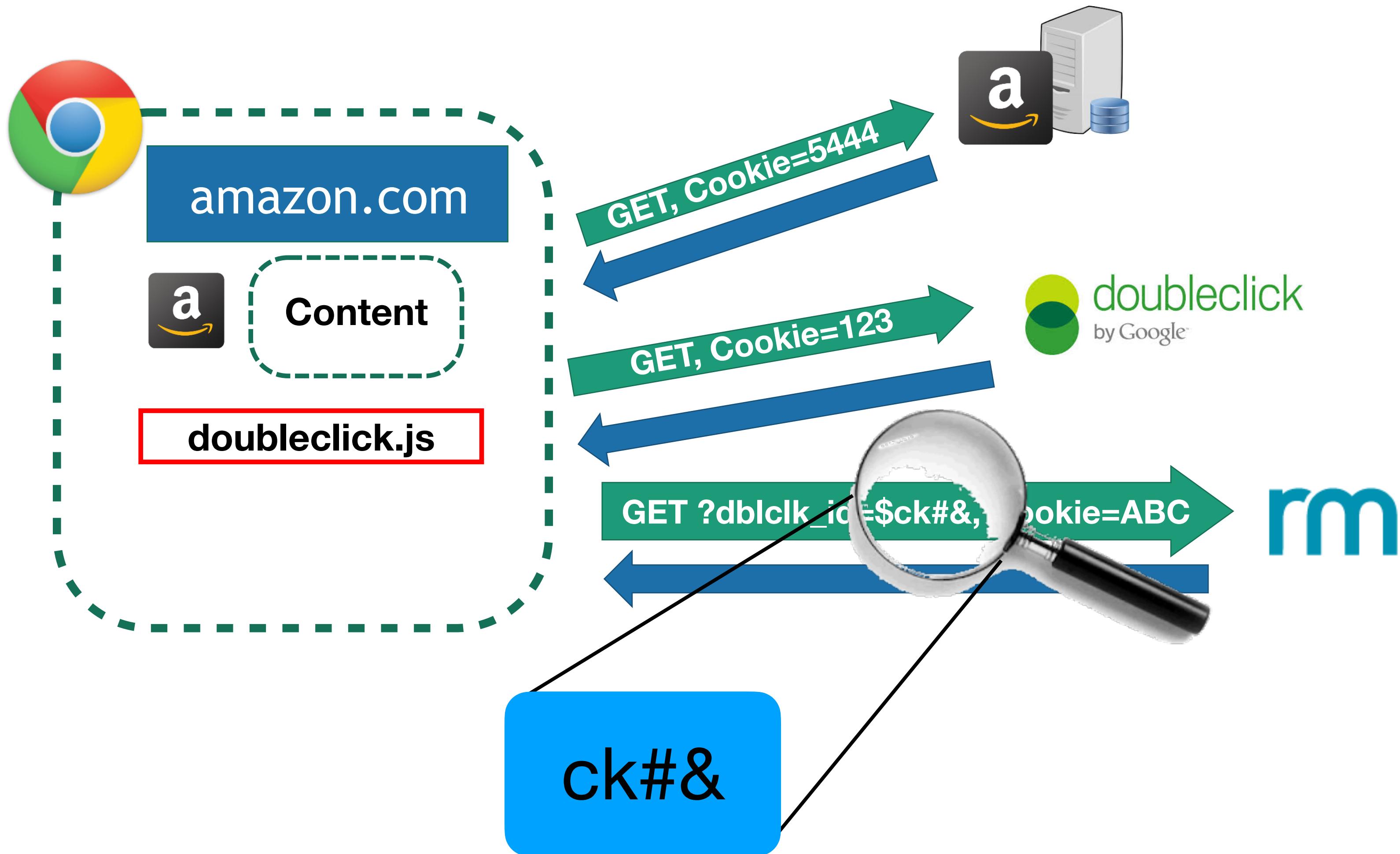
Limitation #2: Obfuscation



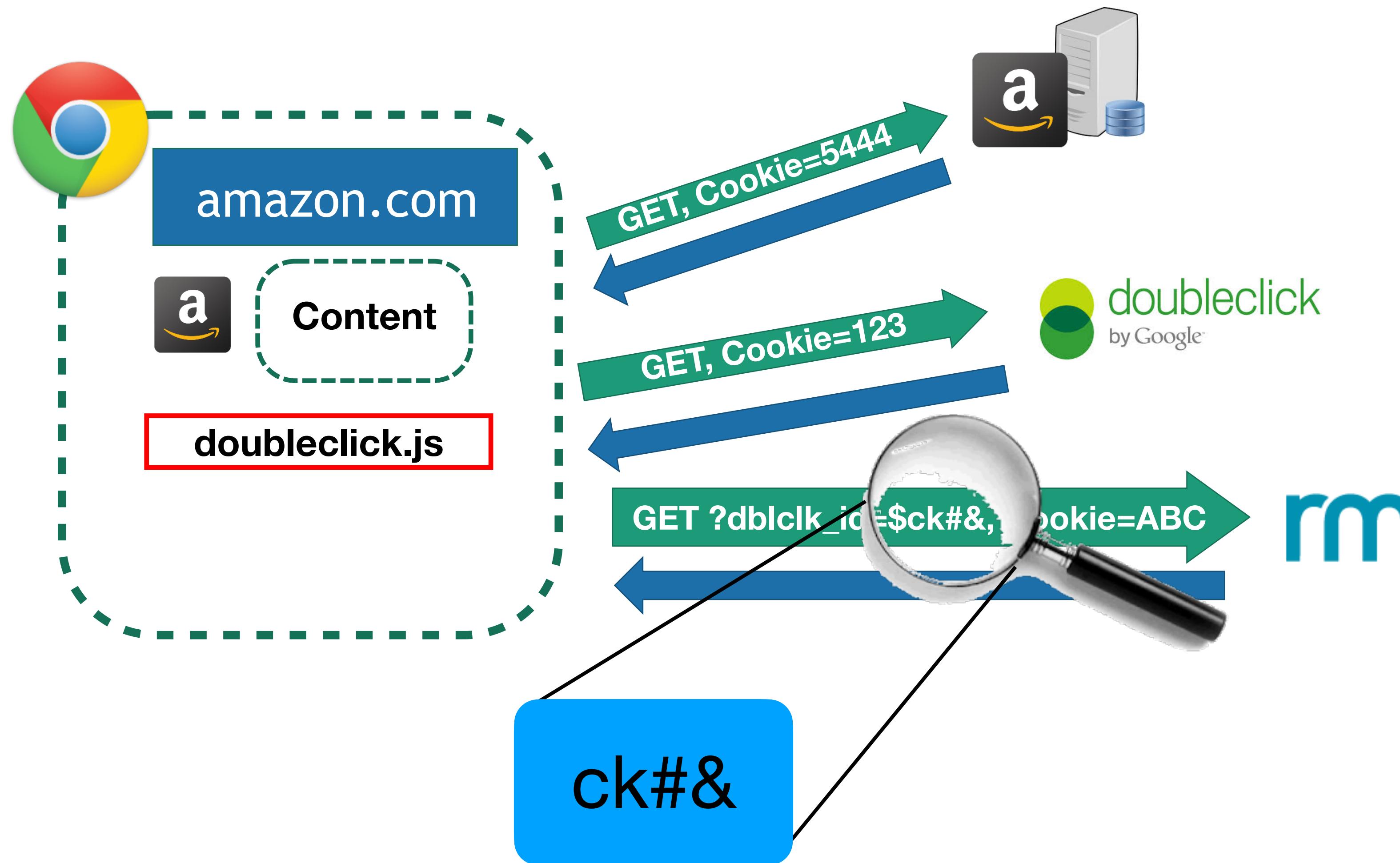
Limitation #2: Obfuscation



Limitation #2: Obfuscation



Limitation #2: Obfuscation

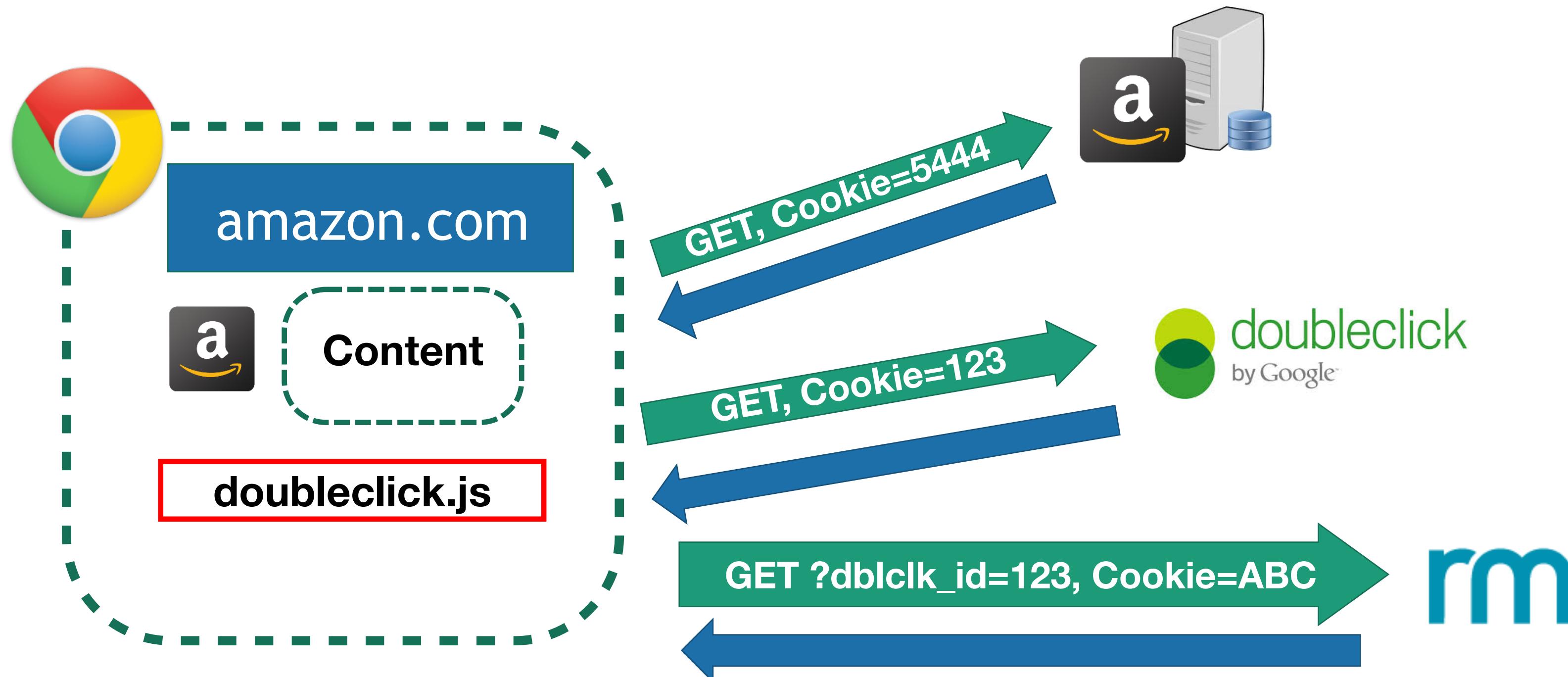


Matching Table

Partner	My ID	PID
DoubleClick	ABC	123

We won't be able to detect Cookie Matching using HTTP headers

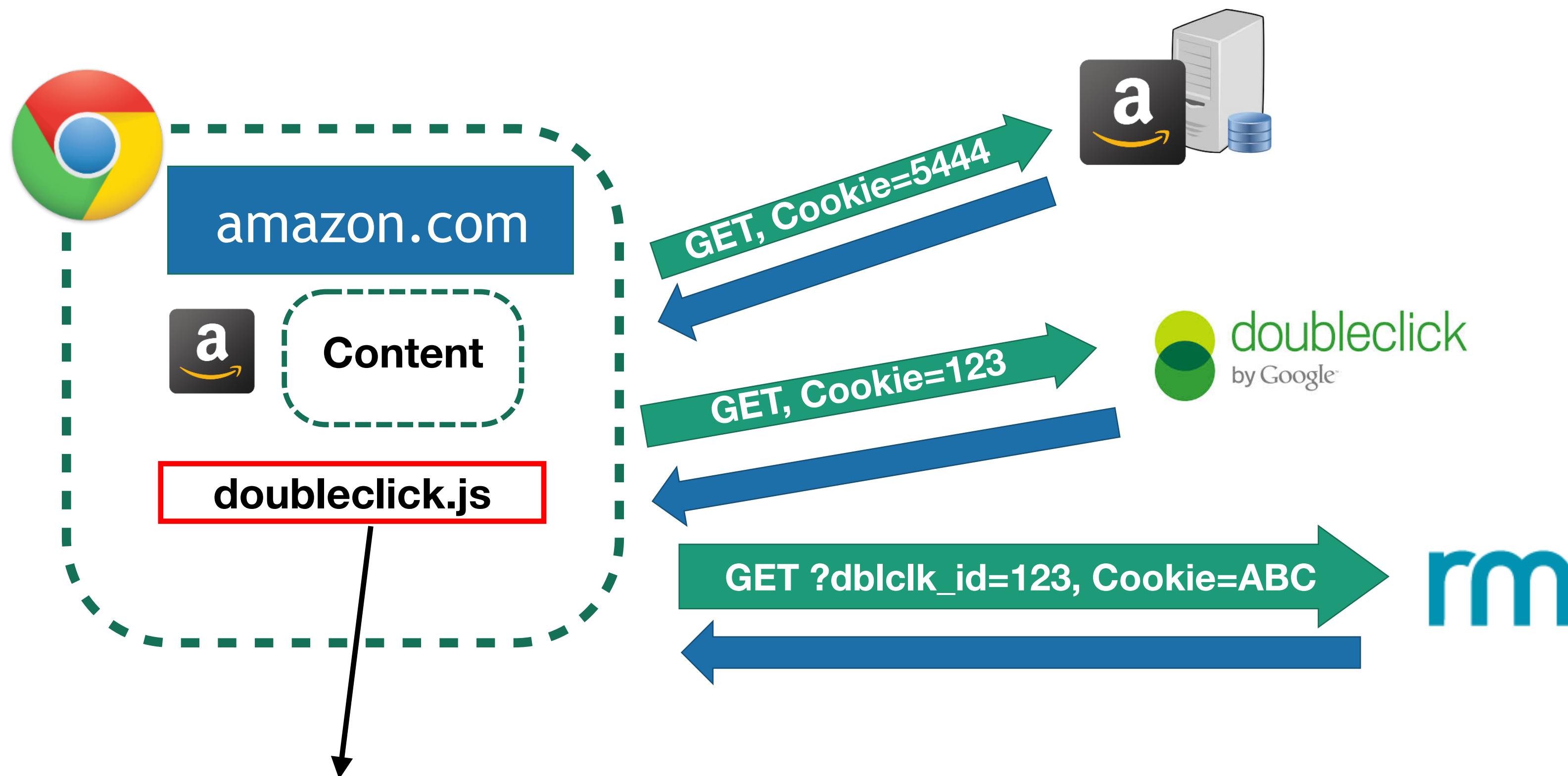
Limitation #3: Attribution of Resources



Matching Table

Partner	My ID	PID
DoubleClick	ABC	123

Limitation #3: Attribution of Resources

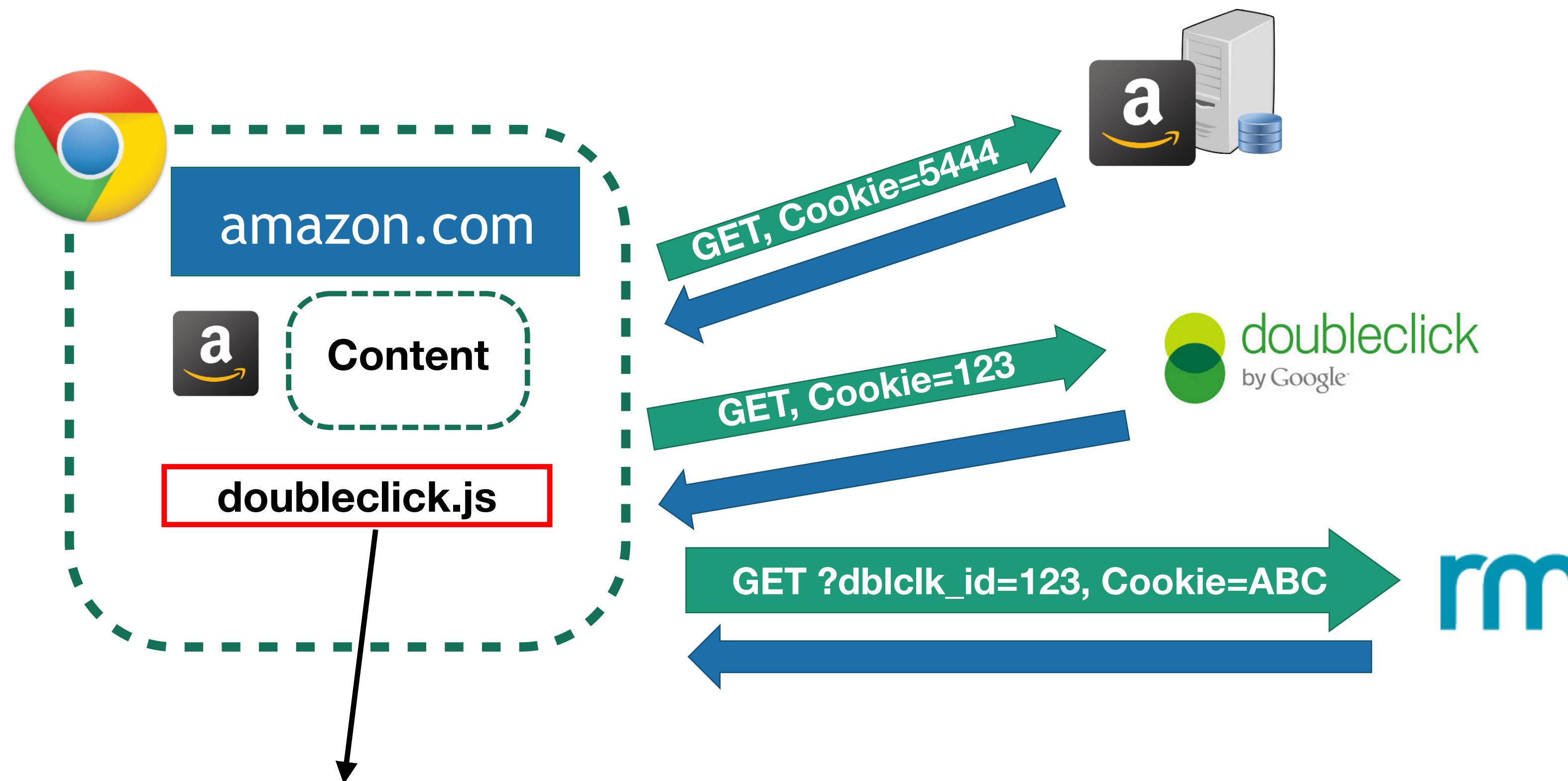


Matching Table

Partner	My ID	PID
DoubleClick	ABC	123

Since JavaScript from DoubleClick is in the context of the main page, the **Referer will be Amazon**

Limitation #3: Attribution of Resources



Matching Table

Partner	My ID	PID
DoubleClick	ABC	123

Since JavaScript from DoubleClick is in the context of the main page, the **Referer will be Amazon**

Referer value in HTTP header can lead to incorrect attribution

A New Approach To Detect Information Sharing



A New Approach To Detect Information Sharing



- 1. Resilient to obfuscation**
 - Should be able to detect matching without relying on content.
- 2. Platform agnostic**
 - Should be able to detect both client-side (browser) and server-side matching.
- 3. Provides strong attribution**
 - Should be able to detect who initiated matching and with whom.

Outline

Background & Motivation

Limitations of Prior Work

Detecting Information Sharing

- Using Retargeted Ads

Propagation of Tracking Data

- Modeling via Graph

Ongoing Work

- Cross-device Tracking
- Automation of Adblocking Filter Lists

Research Plan

Instrumented Chromium for Stronger Attribution

Instrumented Chromium¹ binary that records the provenance of page elements

1: Arshad et al : Include Me Out: In-Browser Detection of Malicious Third-Party Content Inclusions, FC 2016

Instrumented Chromium for Stronger Attribution

Instrumented Chromium¹ binary that records the provenance of page elements

DOM Tree for <http://p.com/index.html>

```
<html>
  <body>
    <script src="a1.com/cookie-match.js"></script>
    <!-- Tracking pixel inserted dynamically
        by cookie-match.js -->
    

    <iframe src="a3.com/banner.html">
      <script src="a4.com/ads.js"></script>
    </iframe>
  </body>
</html>
```

1: Arshad et al : Include Me Out: In-Browser Detection of Malicious Third-Party Content Inclusions, FC 2016

Instrumented Chromium for Stronger Attribution

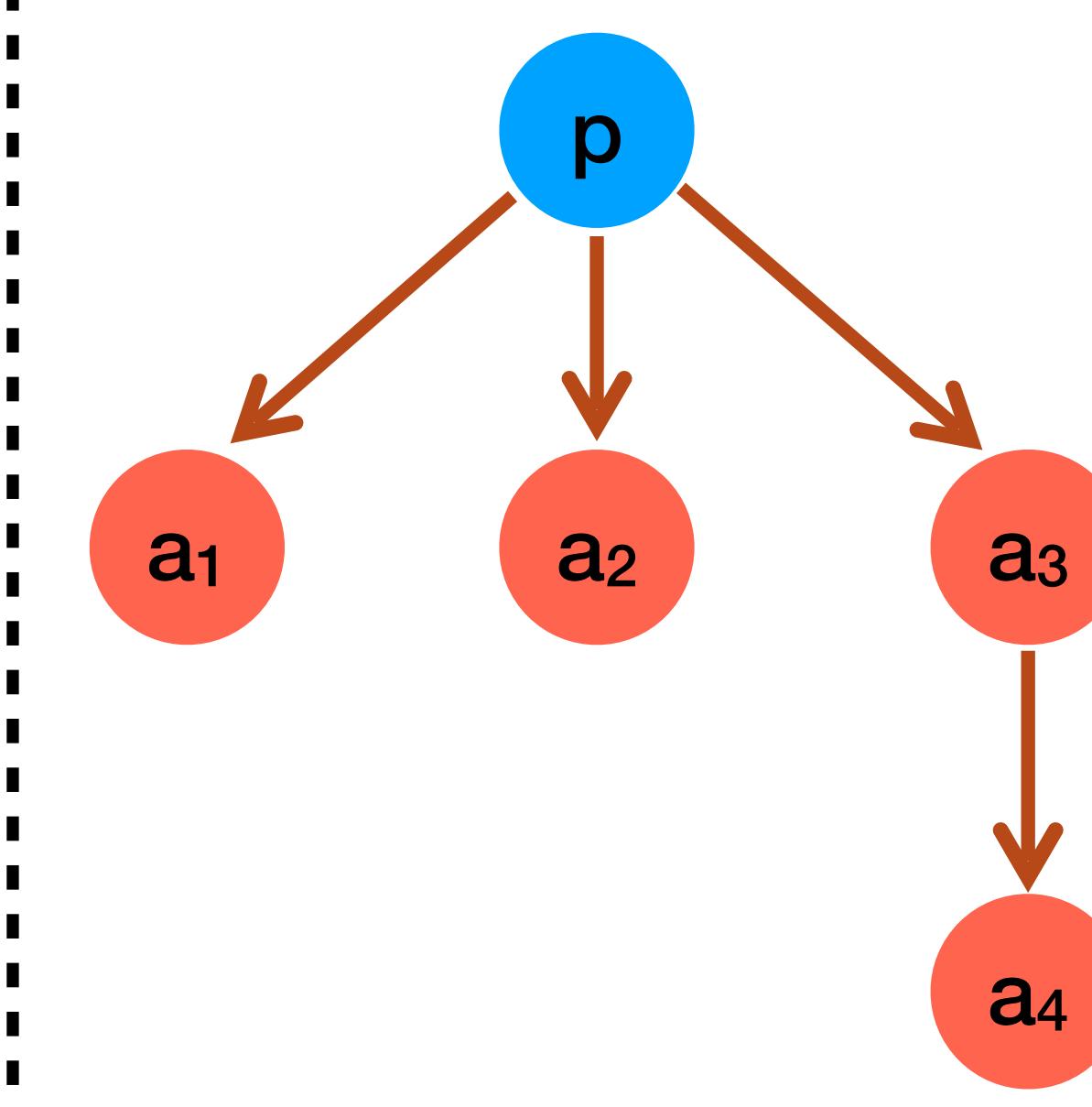
Instrumented Chromium¹ binary that records the provenance of page elements

DOM Tree for <http://p.com/index.html>

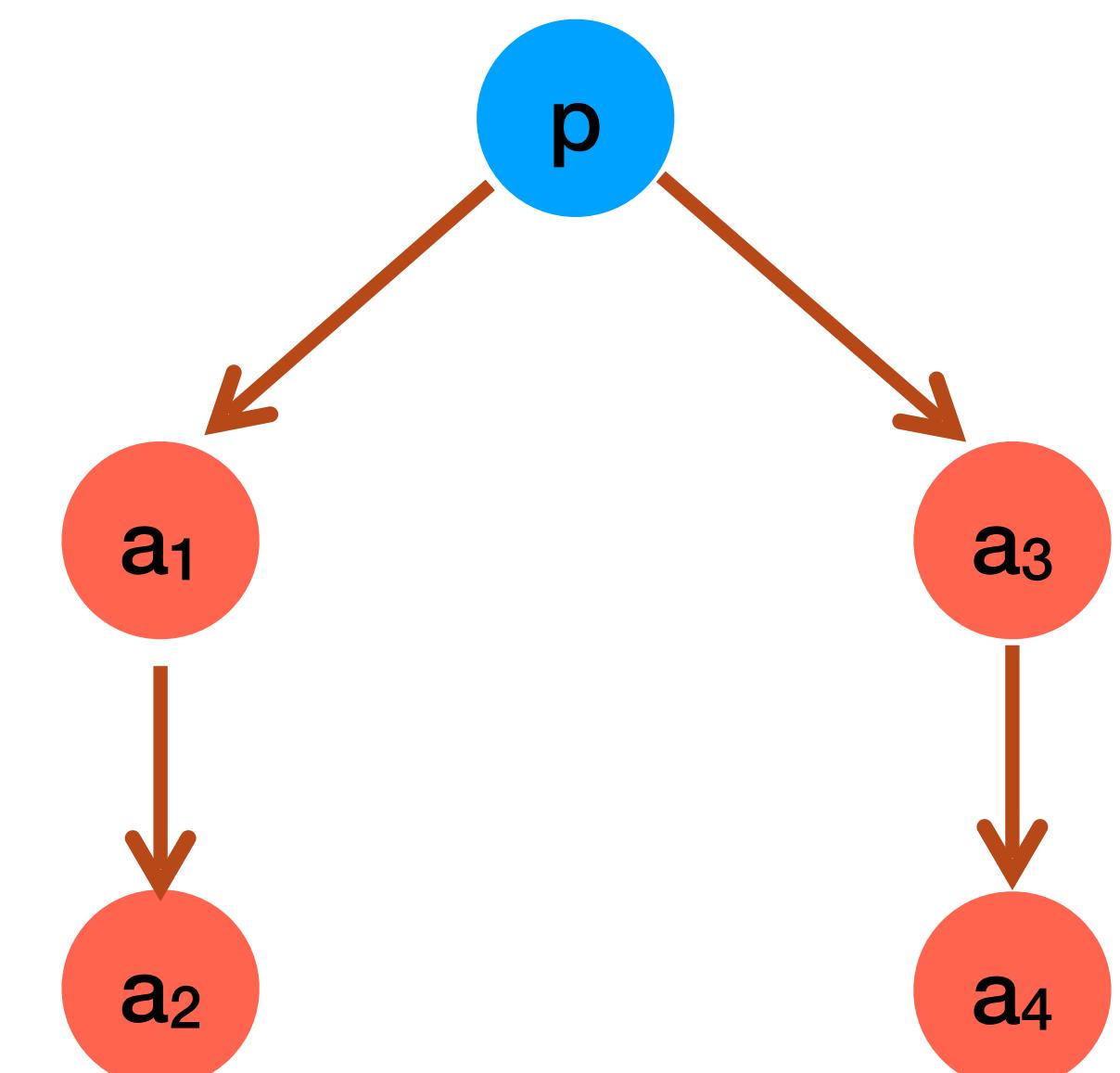
```
<html>
  <body>
    <script src="a1.com/cookie-match.js"></script>
    <!-- Tracking pixel inserted dynamically
        by cookie-match.js -->
    

    <iframe src="a3.com/banner.html">
      <script src="a4.com/ads.js"></script>
    </iframe>
  </body>
</html>
```

Attribution with DOM



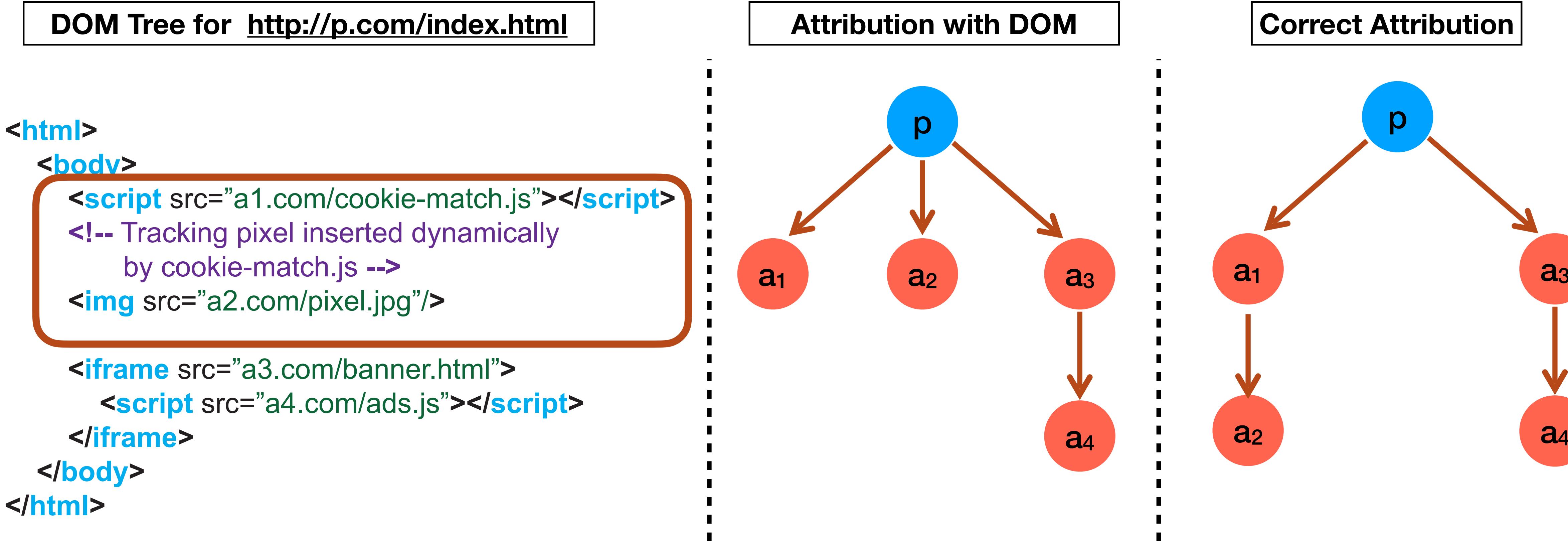
Correct Attribution



1: Arshad et al : Include Me Out: In-Browser Detection of Malicious Third-Party Content Inclusions, FC 2016

Instrumented Chromium for Stronger Attribution

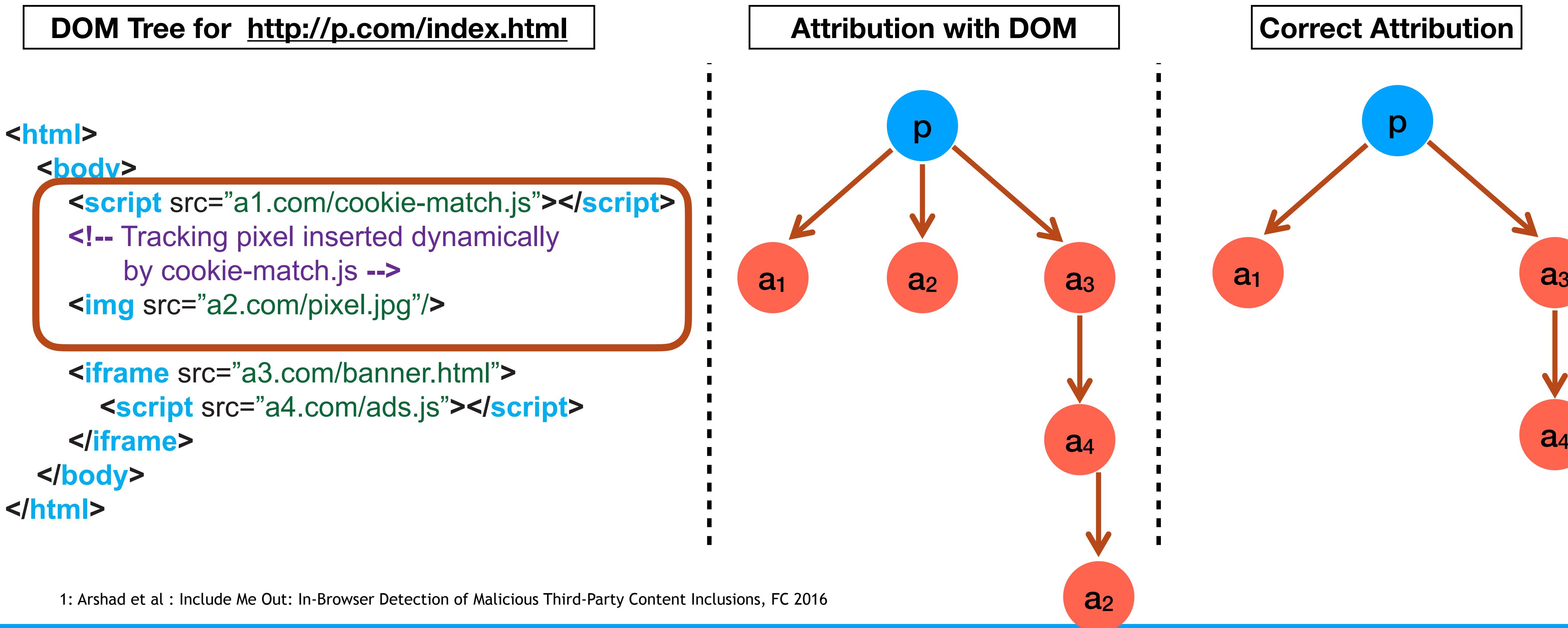
Instrumented Chromium¹ binary that records the provenance of page elements



1: Arshad et al : Include Me Out: In-Browser Detection of Malicious Third-Party Content Inclusions, FC 2016

Instrumented Chromium for Stronger Attribution

Instrumented Chromium¹ binary that records the provenance of page elements



Using Retargeted Ads as a Detection Mechanism

Retargeted ads: The ads which follow you around on the web

- Highly targeted (specific) ads
- Very expensive: around \$1 per impression vs. >\$0.01 for contextual ads

Using Retargeted Ads as a Detection Mechanism

Retargeted ads: The ads which follow you around on the web

- Highly targeted (specific) ads
- Very expensive: around \$1 per impression vs. >\$0.01 for contextual ads



Cisco-Linksys AE1000 High-Performance Wireless-N Adapter
by Linksys
 207 customer reviews | 10 answered questions
Price: **\$15.99** 
Only 1 left in stock.
Want it Tuesday, June 14? Order within 33 hrs 50 mins and choose One-Day Shipping at checkout. [Details](#)
Sold by Home Sweet Home Direct and Fulfilled by Amazon.
Eligible for  donation.

 Want to hire a computer technician?
Buy professional computer technician services directly on Amazon. Backed by our Happiness Guarantee.
[Learn more](#)

• Networking Equipment Features: WEP Security, WPA Security, Easy Setup, WPA2



Key Insight: Because retargeted ads are so specific, they can be used to conduct controlled experiments

- Information must be shared between ad networks to serve retargeted ads

Retargeted Ads are Very Specific (A Simple Experiment)

1)

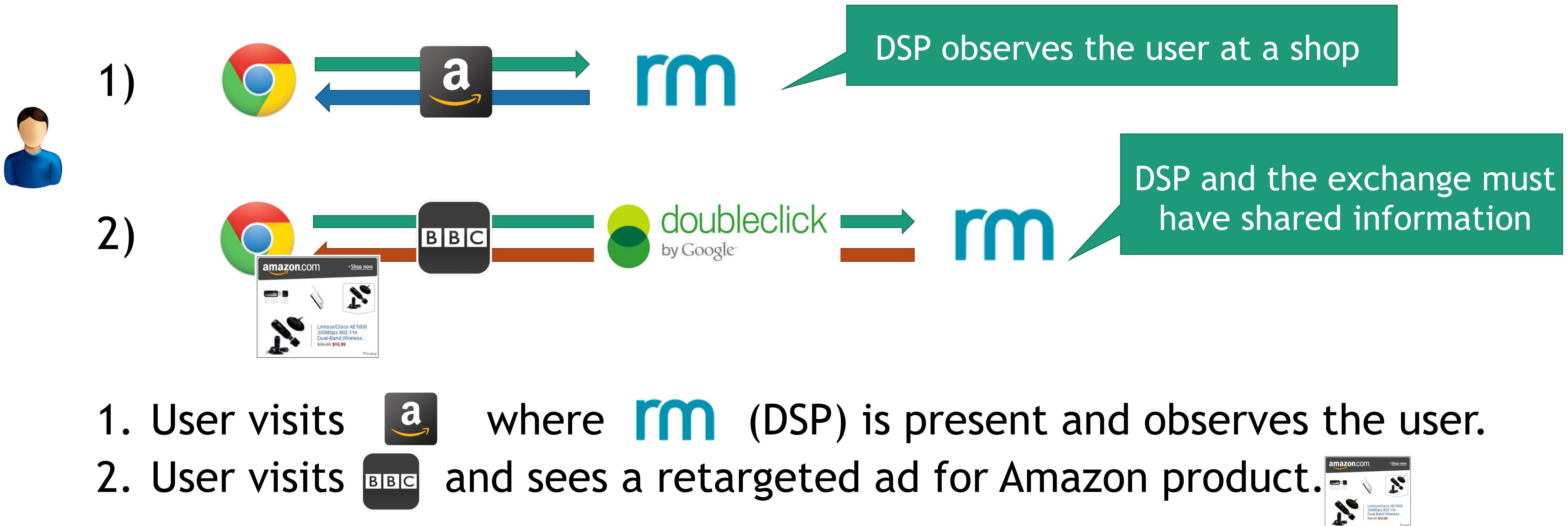


Retargeted Ads are Very Specific (A Simple Experiment)

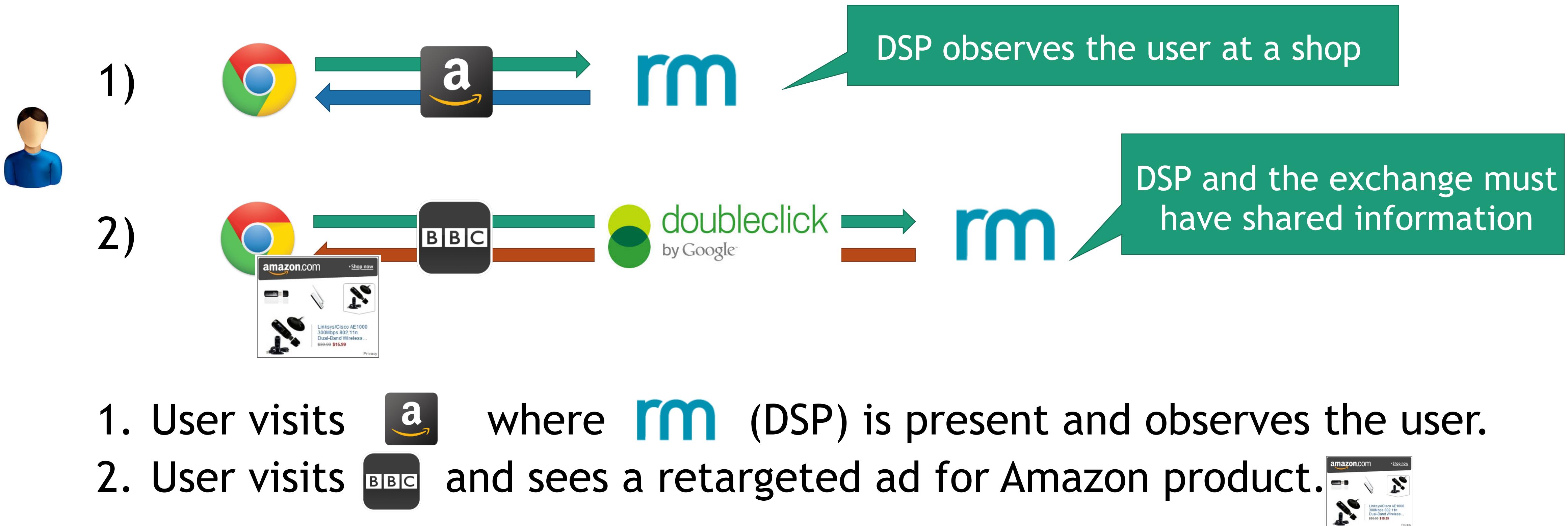


1. User visits where (DSP) is present and observes the user.

Retargeted Ads are Very Specific (A Simple Experiment)



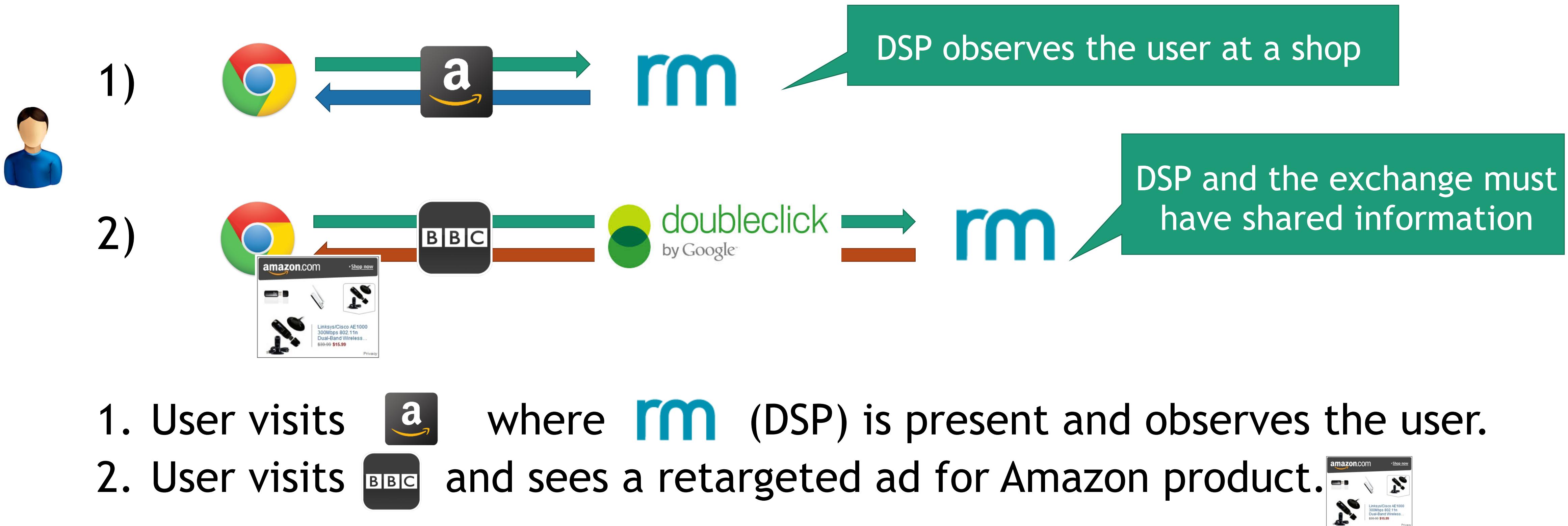
Retargeted Ads are Very Specific (A Simple Experiment)



How does **rm** identify the user on **BBC**, to show a specific ad?

- Retargets are too expensive to be served to random, unknown users

Retargeted Ads are Very Specific (A Simple Experiment)



How does identify the user on , to show a specific ad?

- Retargets are too expensive to be served to random, unknown users
- This proves a causal flow of information from (exchange) → (DSP)
- **Note:** We do not rely on HTTP headers to determine sharing.

Limitations Addressed

Limitations Addressed

Strong Attribution

Content Agnostic

Platform Agnostic

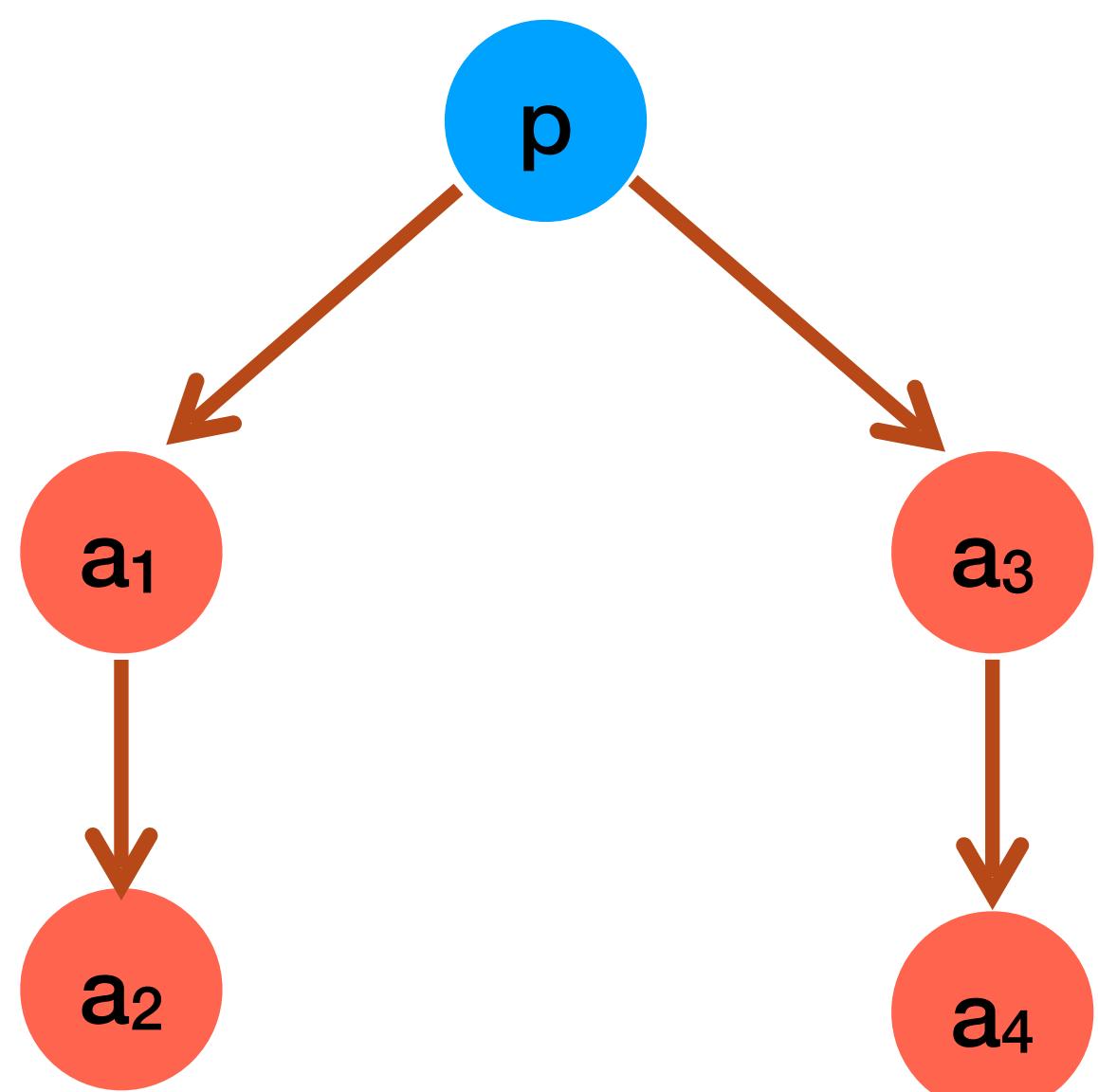
Limitations Addressed

Strong Attribution

Content Agnostic

Platform Agnostic

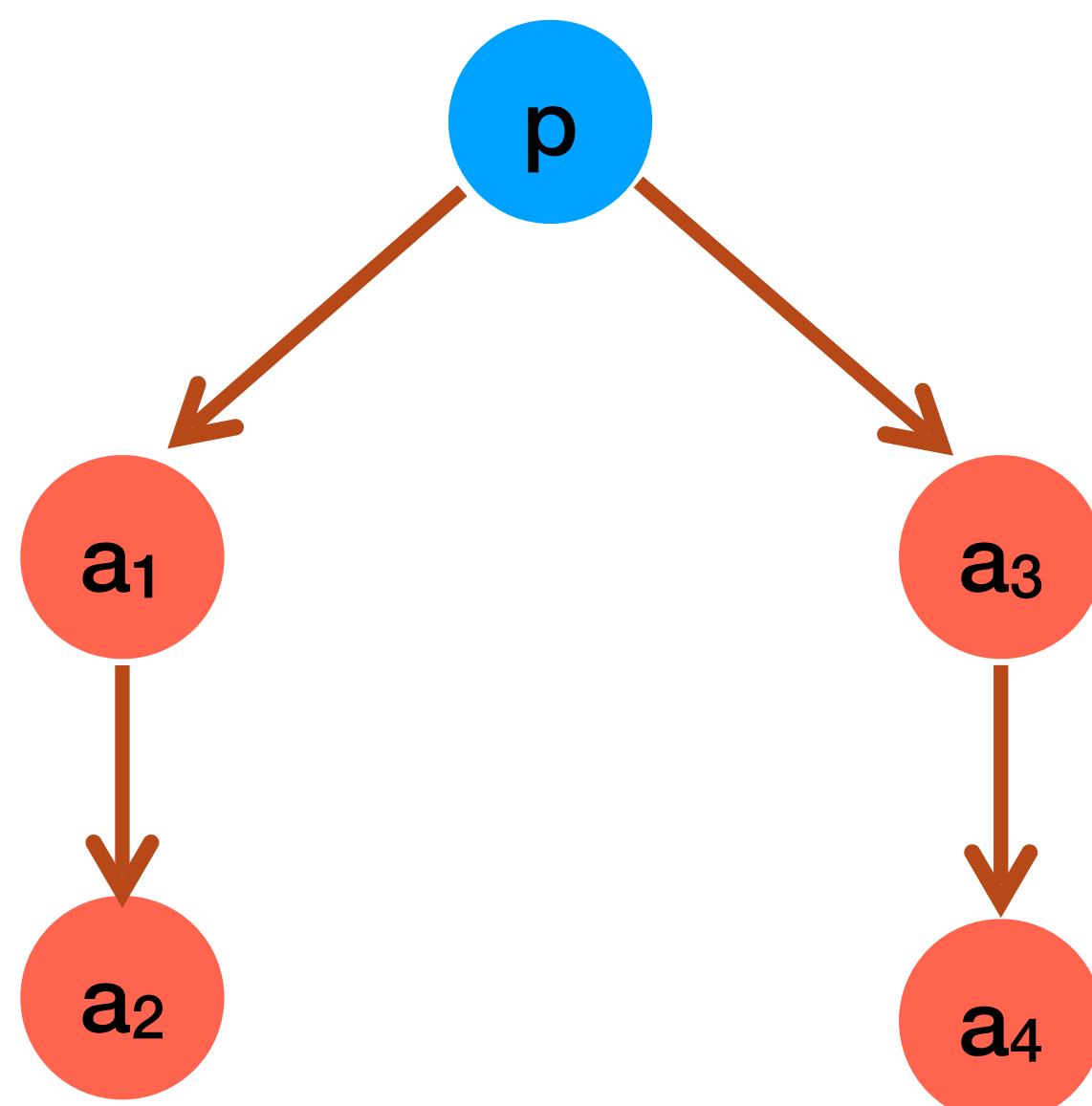
Inclusion Chains



Limitations Addressed

Strong Attribution

Inclusion Chains



Content Agnostic



Cisco-Linksys AE1000 High-Performance Wireless-N Adapter

by [Linksys](#)

 207 customer reviews | 10 answered questions

Price: **\$15.99** 

Only 1 left in stock.

Want it Tuesday, June 14? Order within **33 hrs 50 mins** and choose **One-Day Shipping** at checkout. [Details](#)

Sold by [Home Sweet Home Direct](#) and [Fulfilled by Amazon](#).

Eligible for [amazon smile](#) donation.

 Want to hire a computer technician?

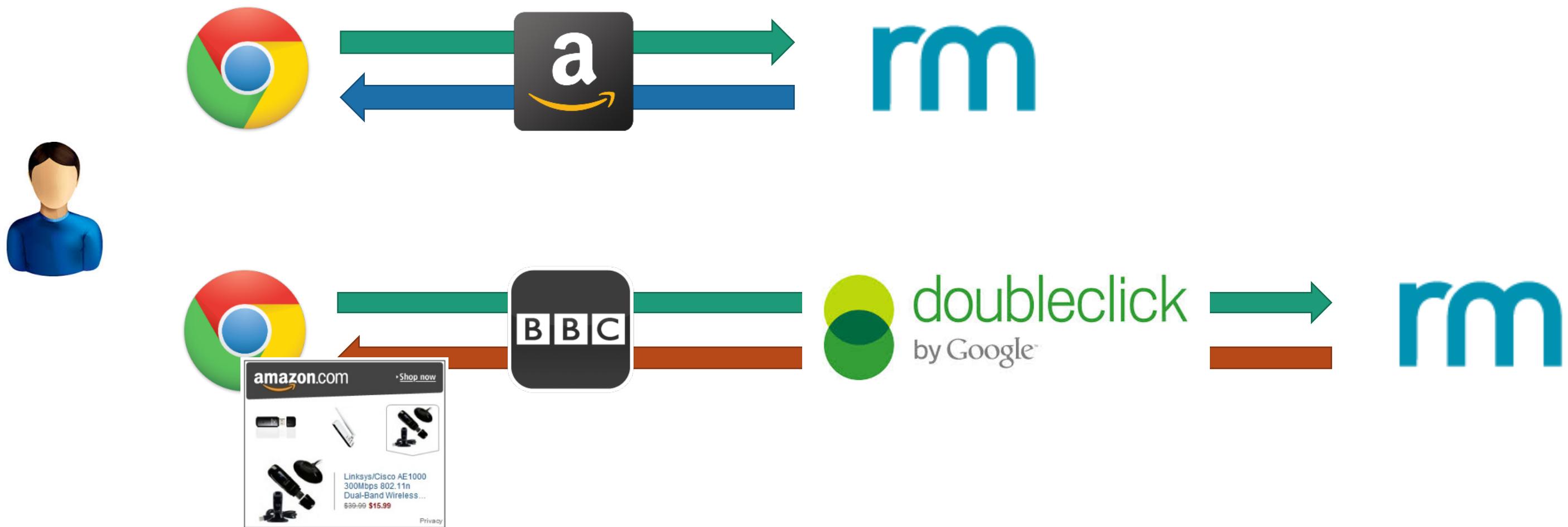
Buy professional computer technician services directly on Amazon. Backed by our Happiness Guarantee.

[Learn more](#)

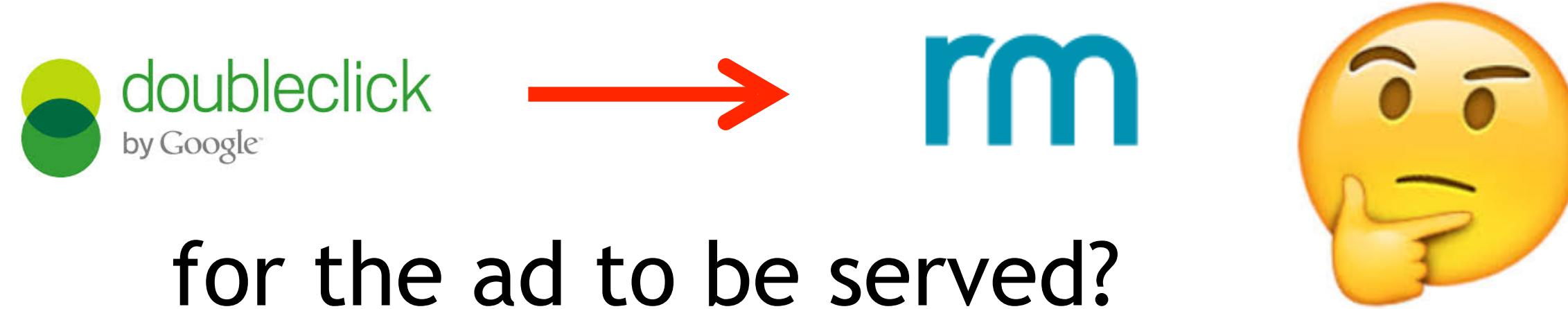
- Networking Equipment Features: WEP Security, WPA Security, Easy Setup, WPA2



The Caveat

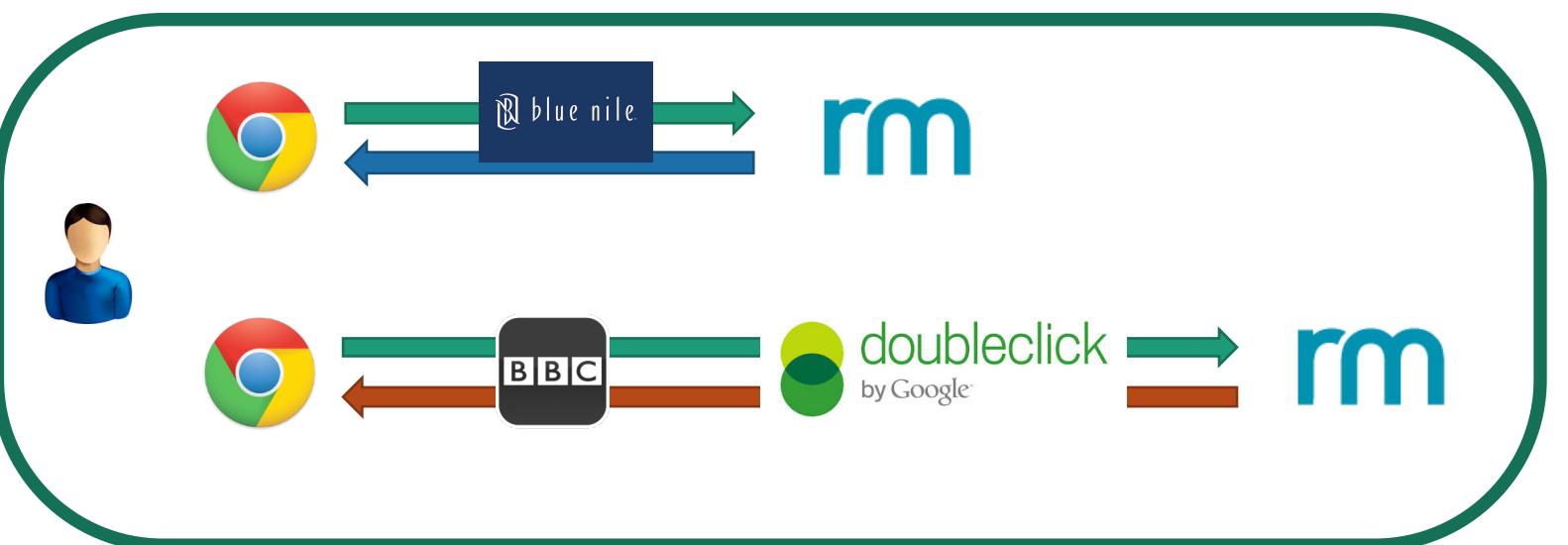
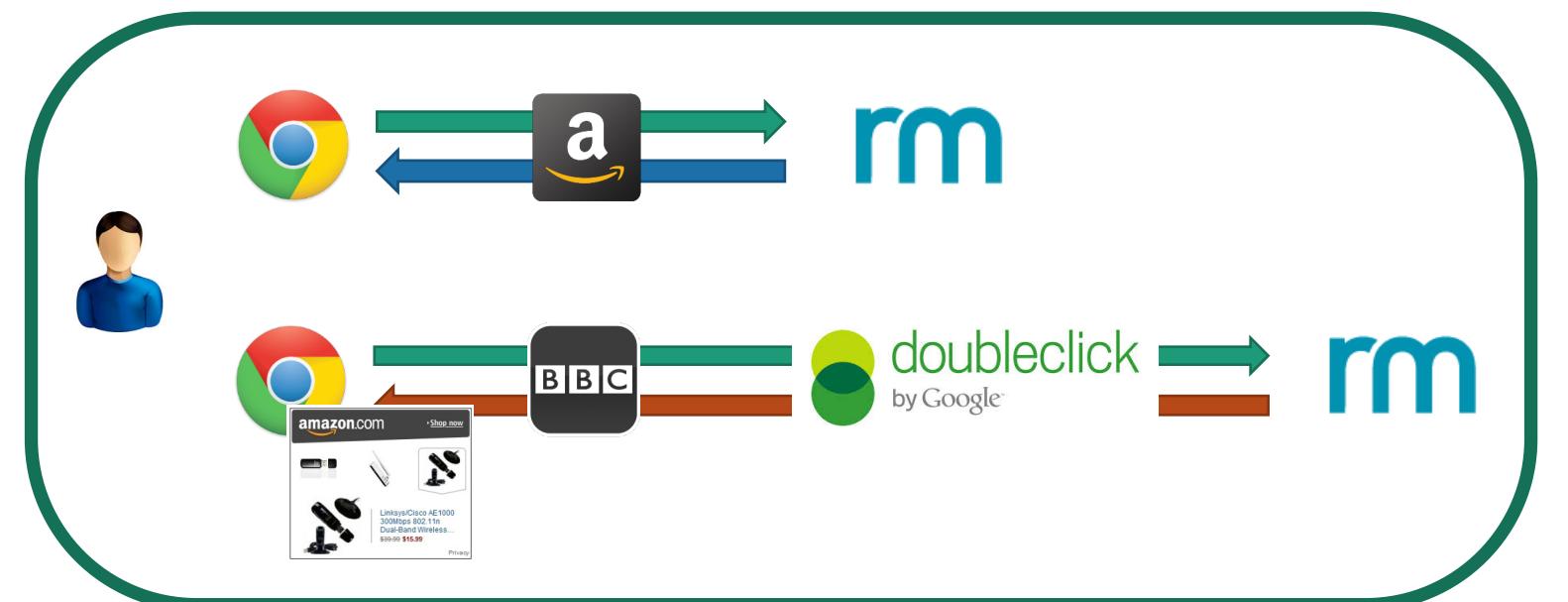


How can we be certain that there is a flow of tracking information from

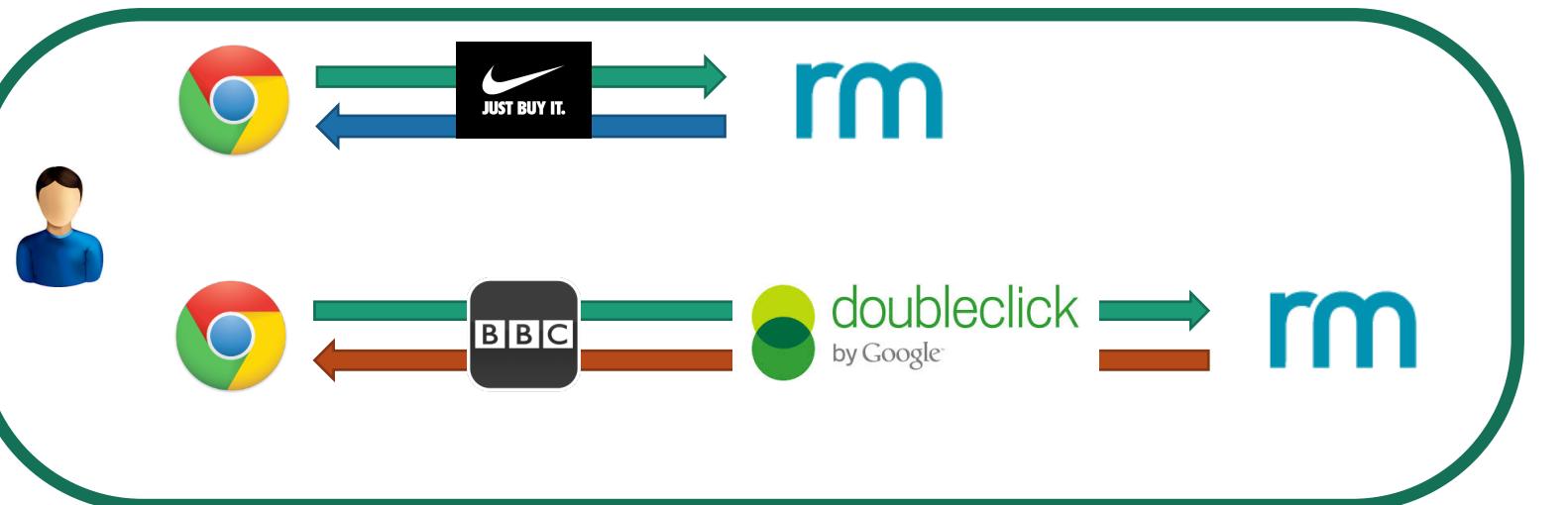


It is possible that amazon is running a very aggressive campaign with a huge budget

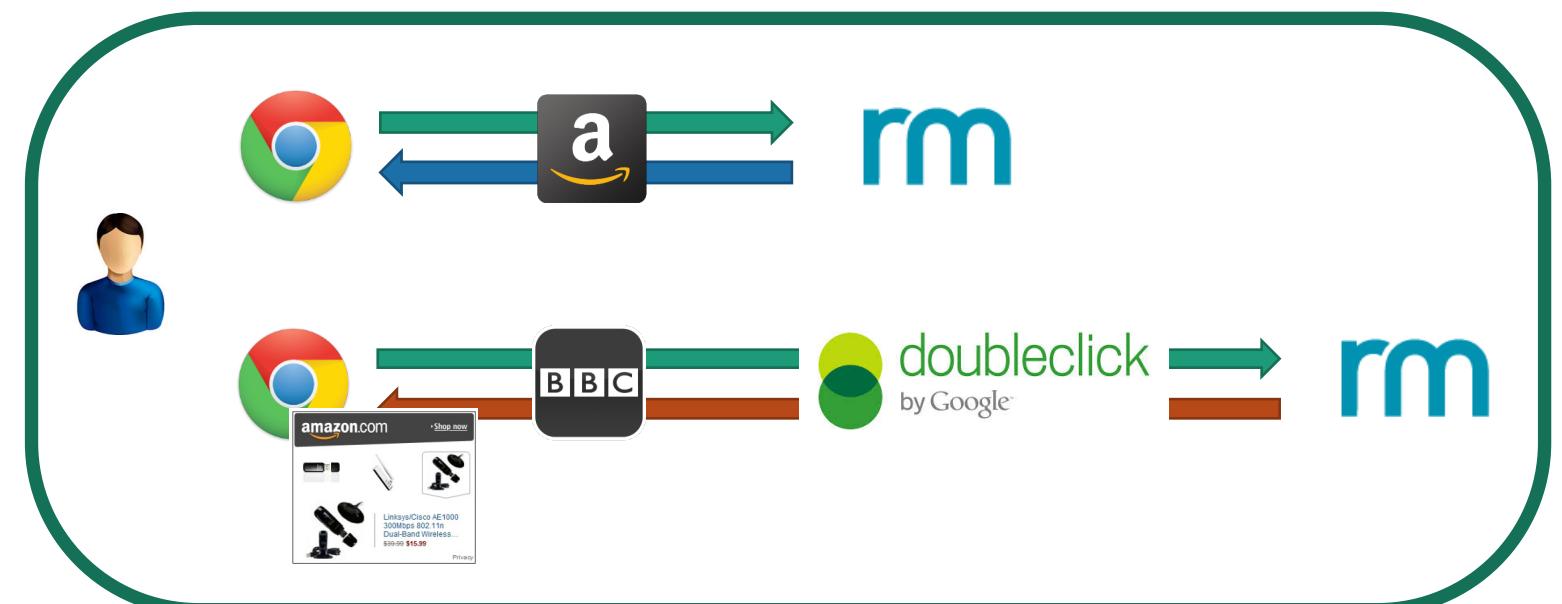
Controlled Experiments



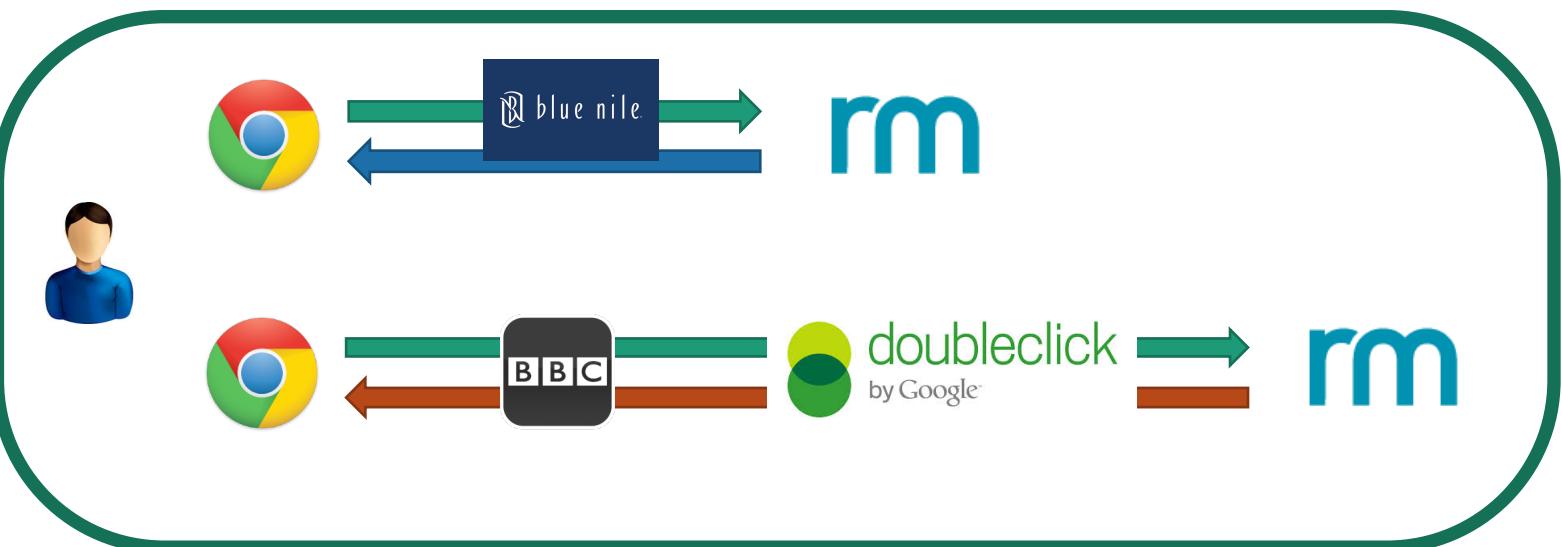
-



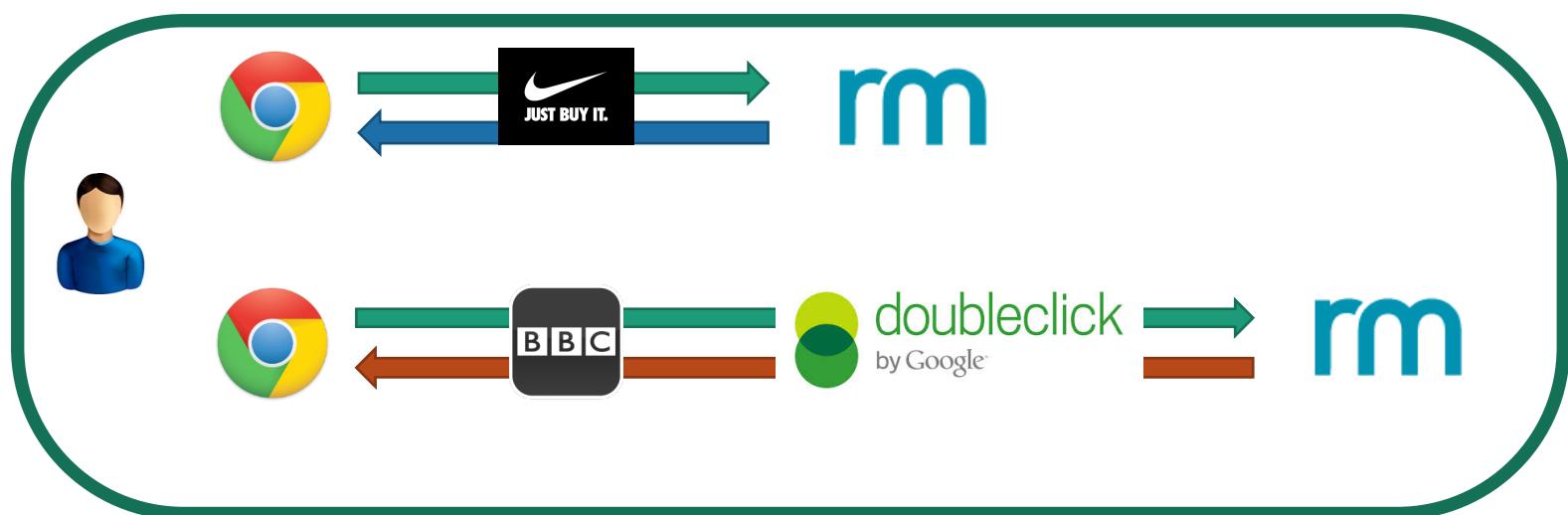
Controlled Experiments



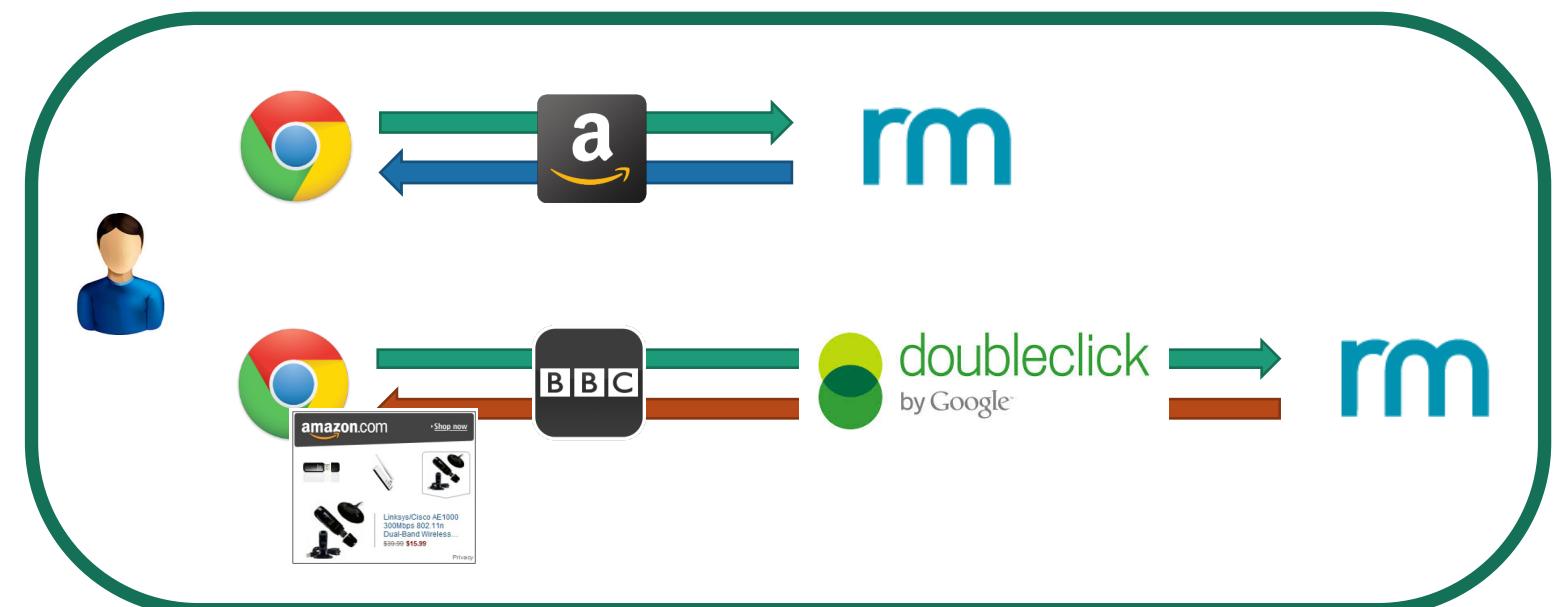
Now imagine that we have 100 users.



.

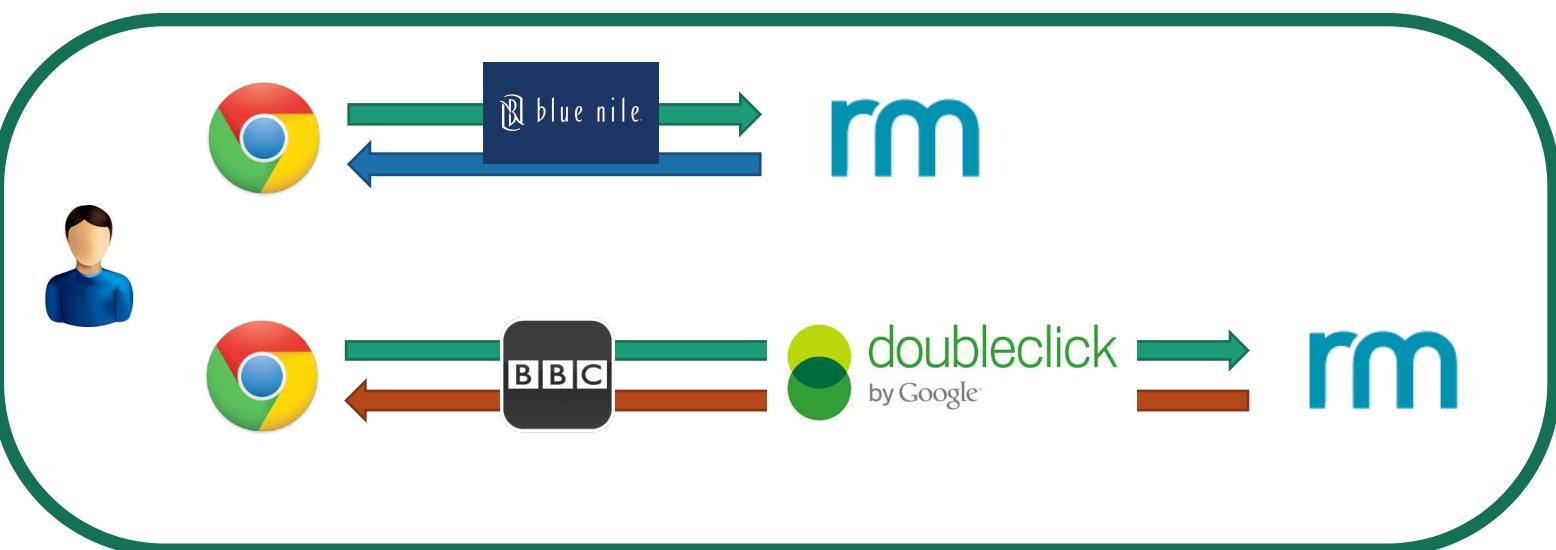


Controlled Experiments

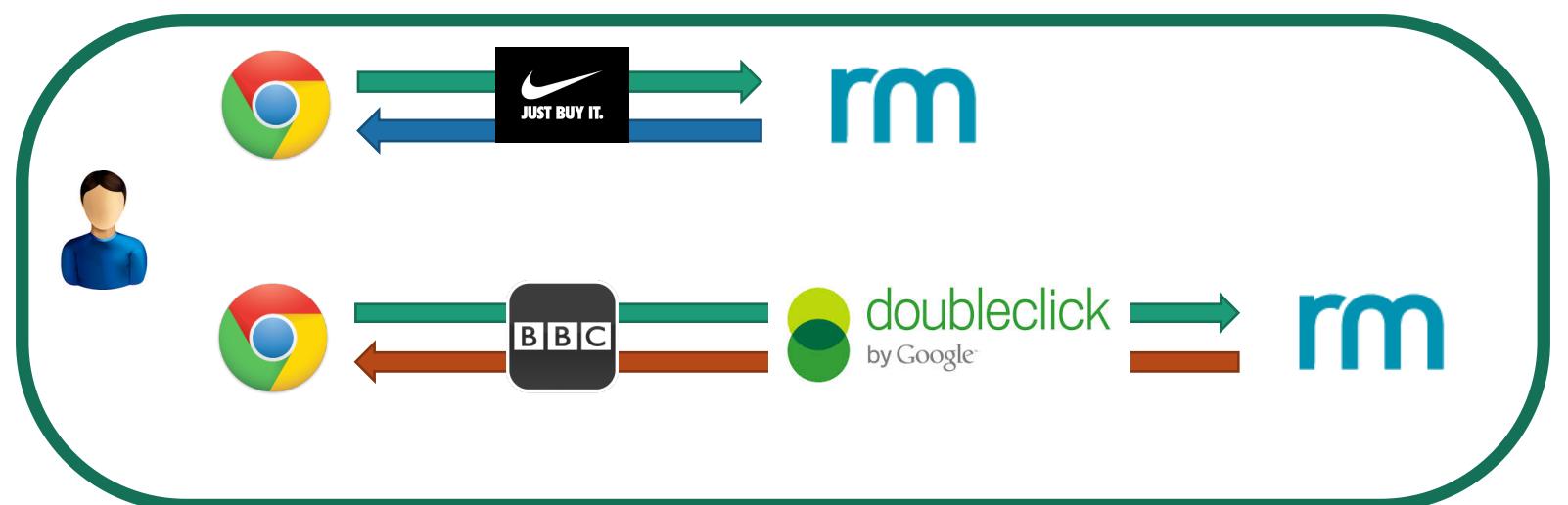


Now imagine that we have 100 users.

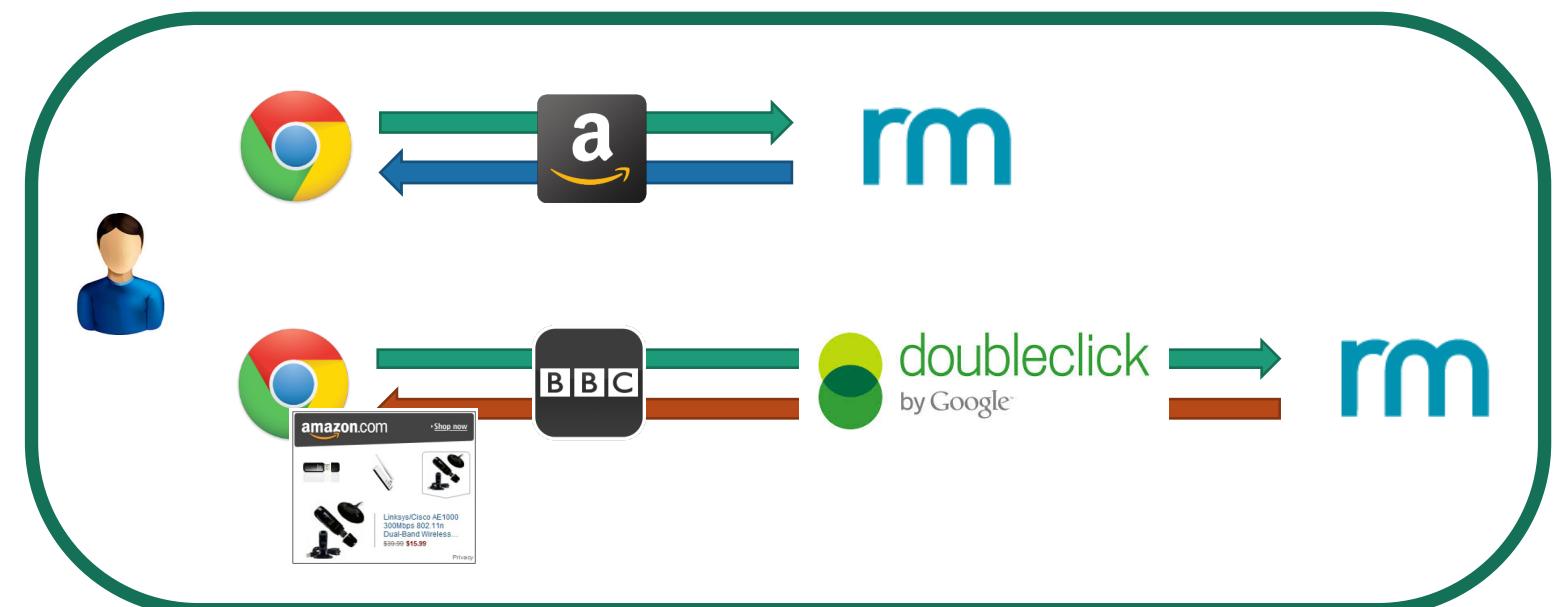
- Only the 1st user visits amazon.
- ... and eventually sees an ad for amazon product.



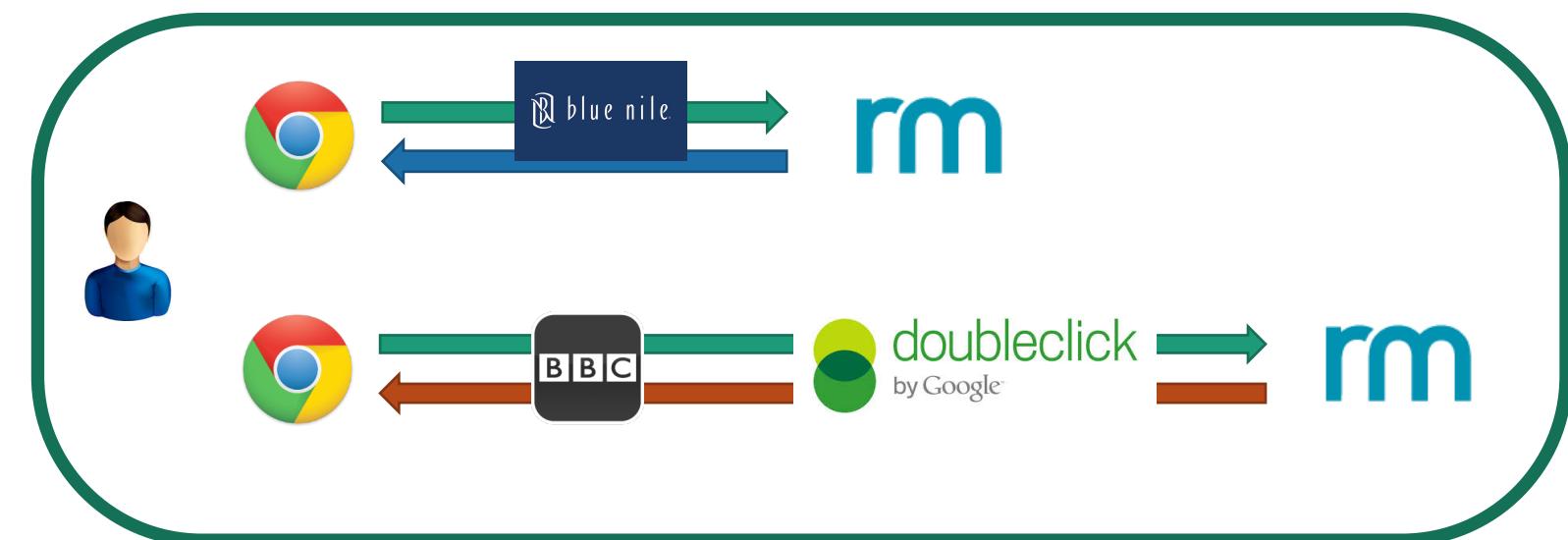
.



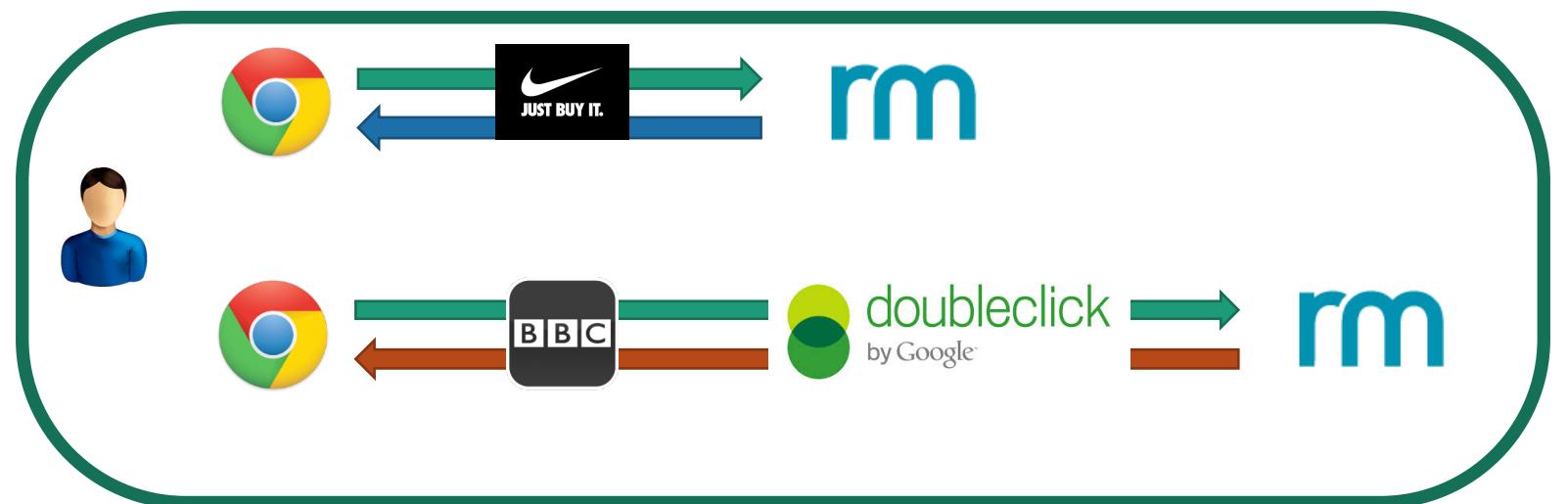
Controlled Experiments



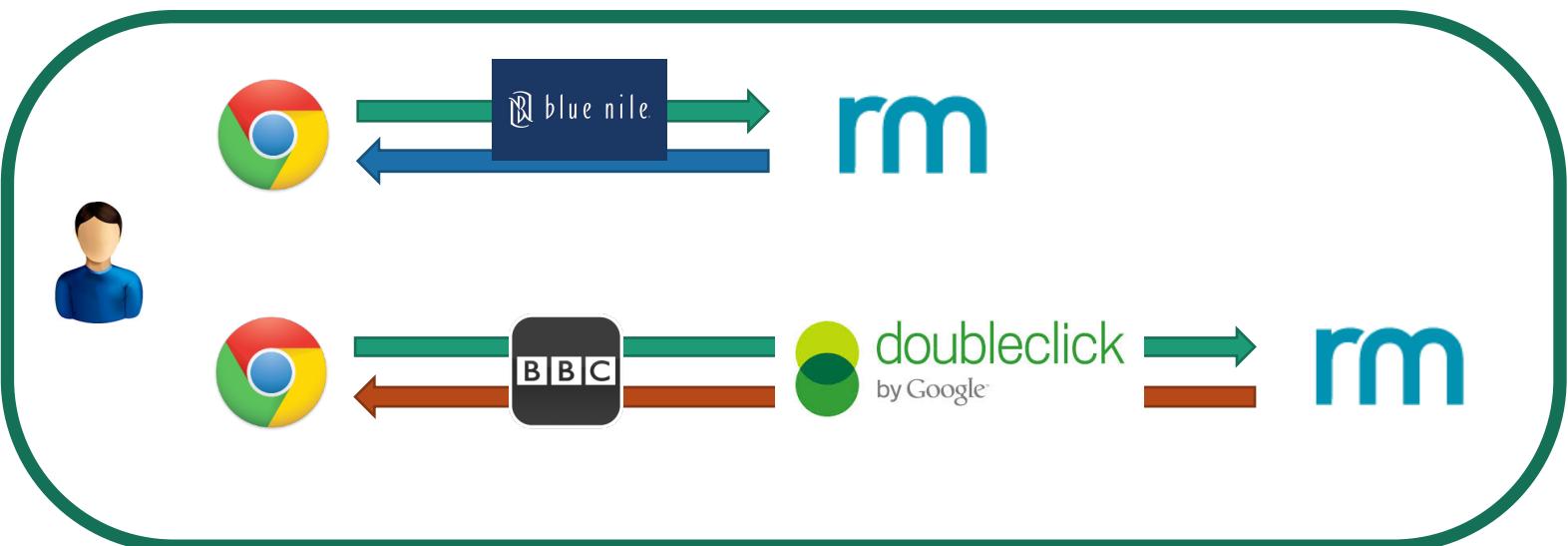
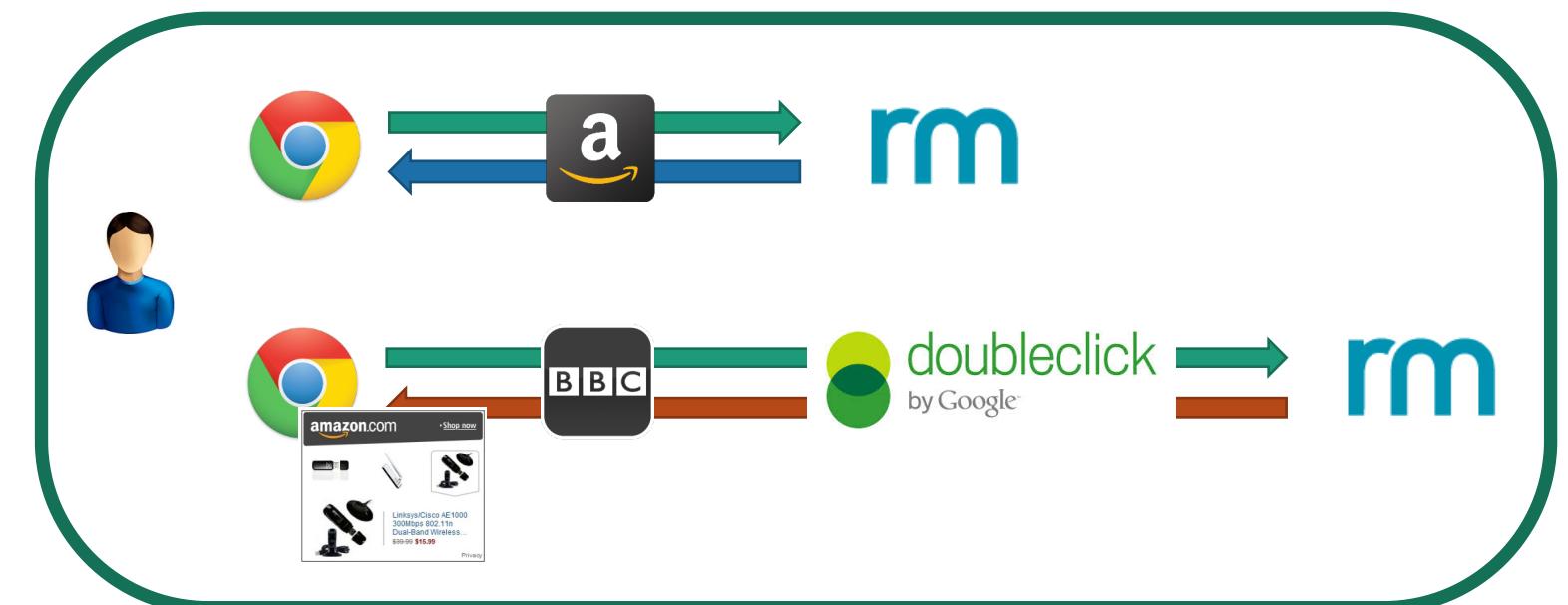
Now imagine that we have 100 users.



- Only the 1st user visits amazon.
- ... and eventually sees an ad for amazon product.
- Now we are much more confident about information sharing between DoubleClick and RightMedia.



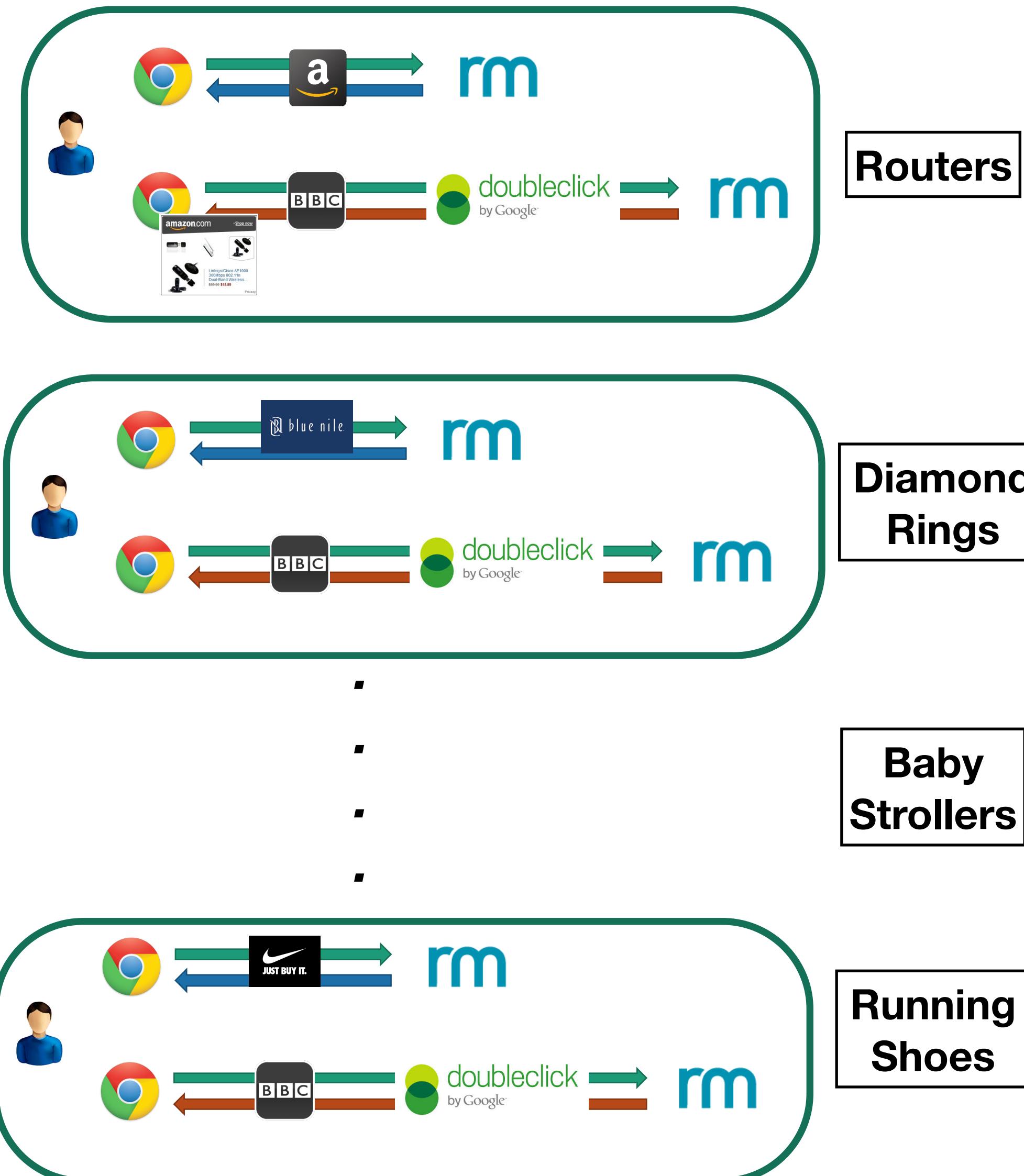
Experiment Setup & Data Collection



-



Experiment Setup & Data Collection



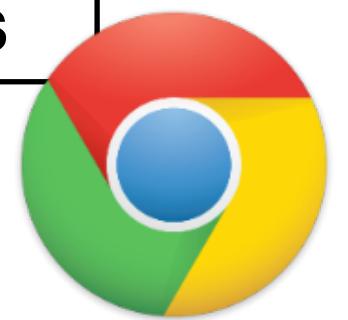
We set up 90 personas (interest profiles)

- Each Persona is very specific
- We select 10 websites per persona
- ... and 10 products per website

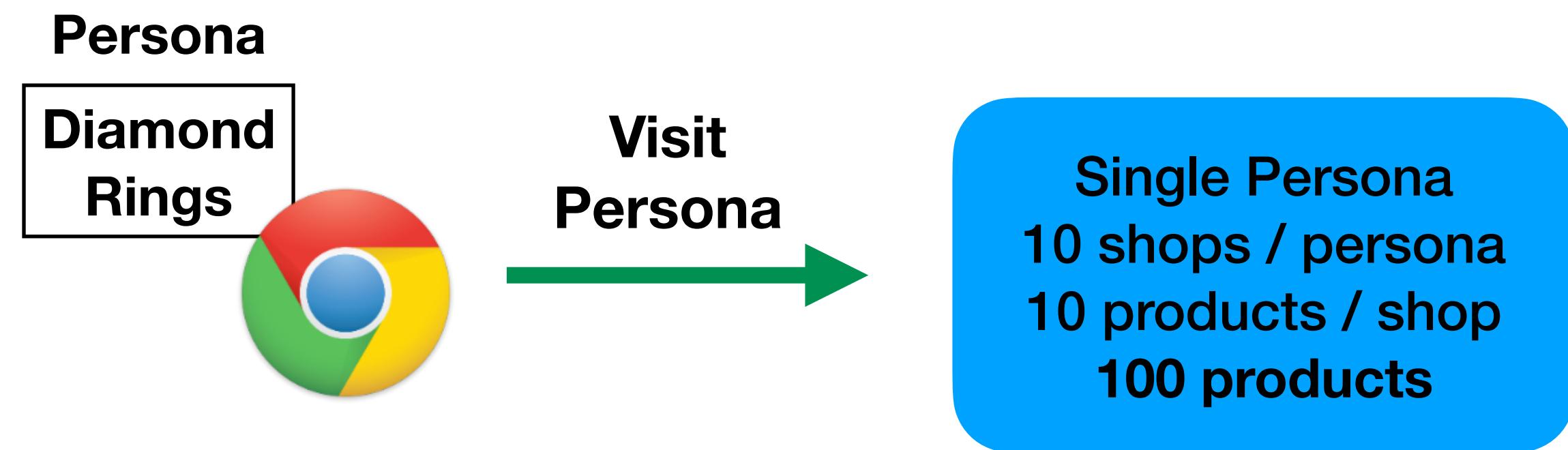
Experiment Setup & Data Collection

Persona

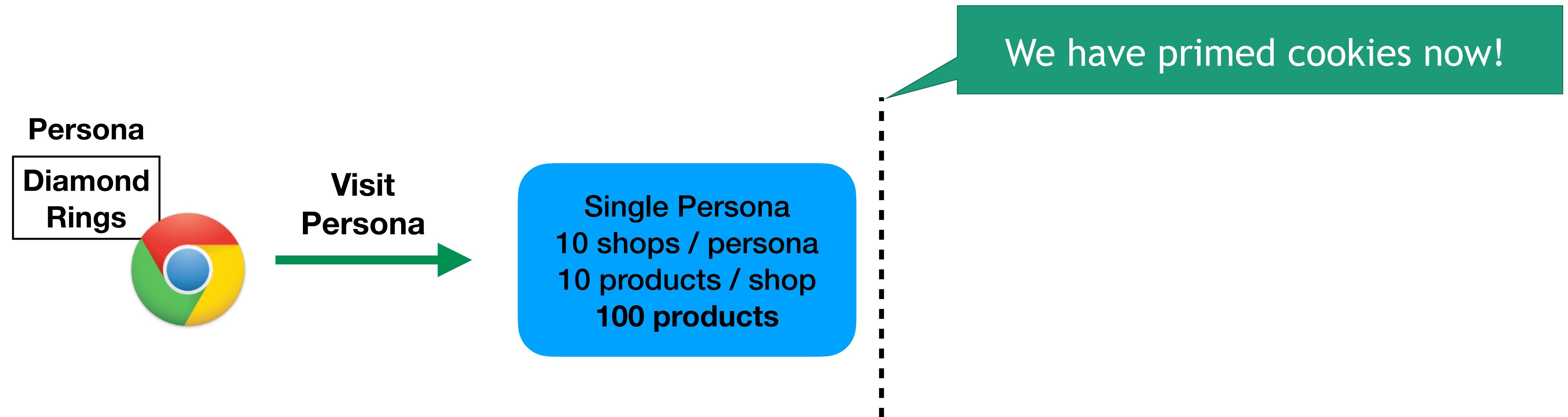
**Diamond
Rings**



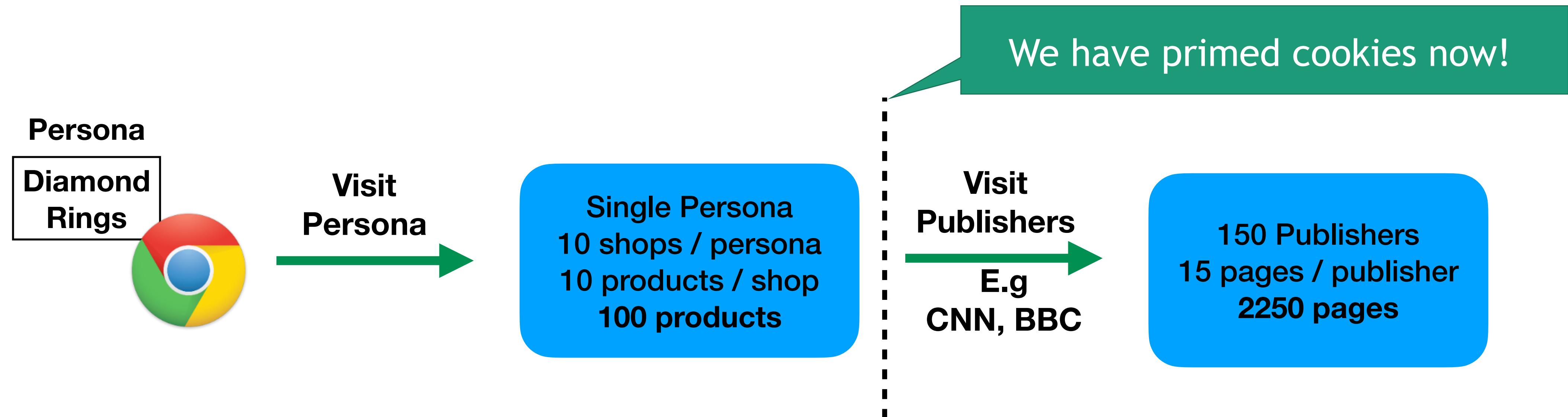
Experiment Setup & Data Collection



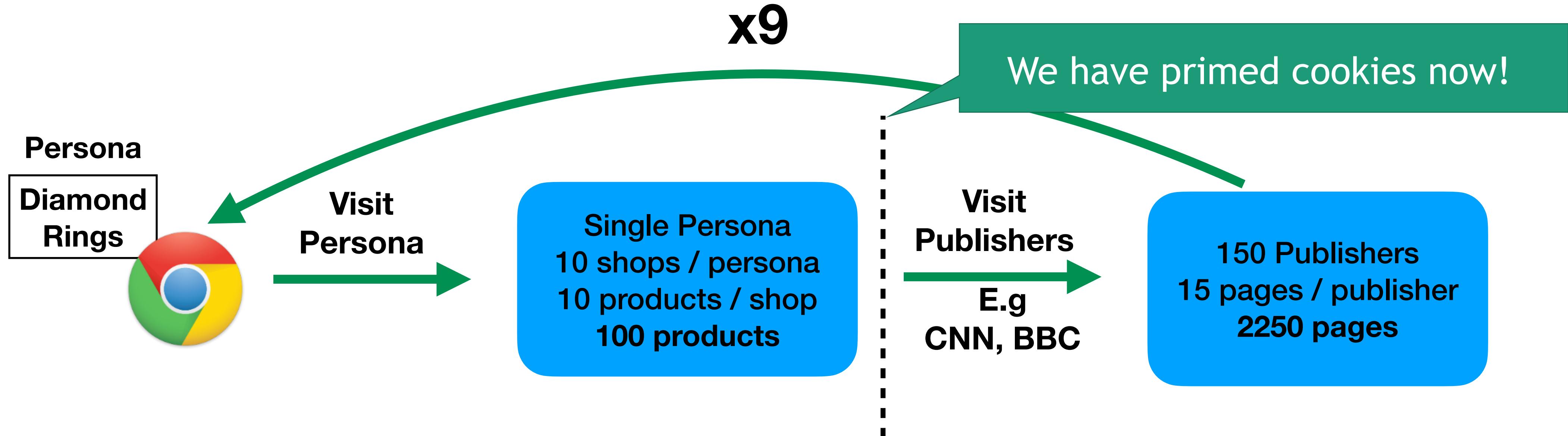
Experiment Setup & Data Collection



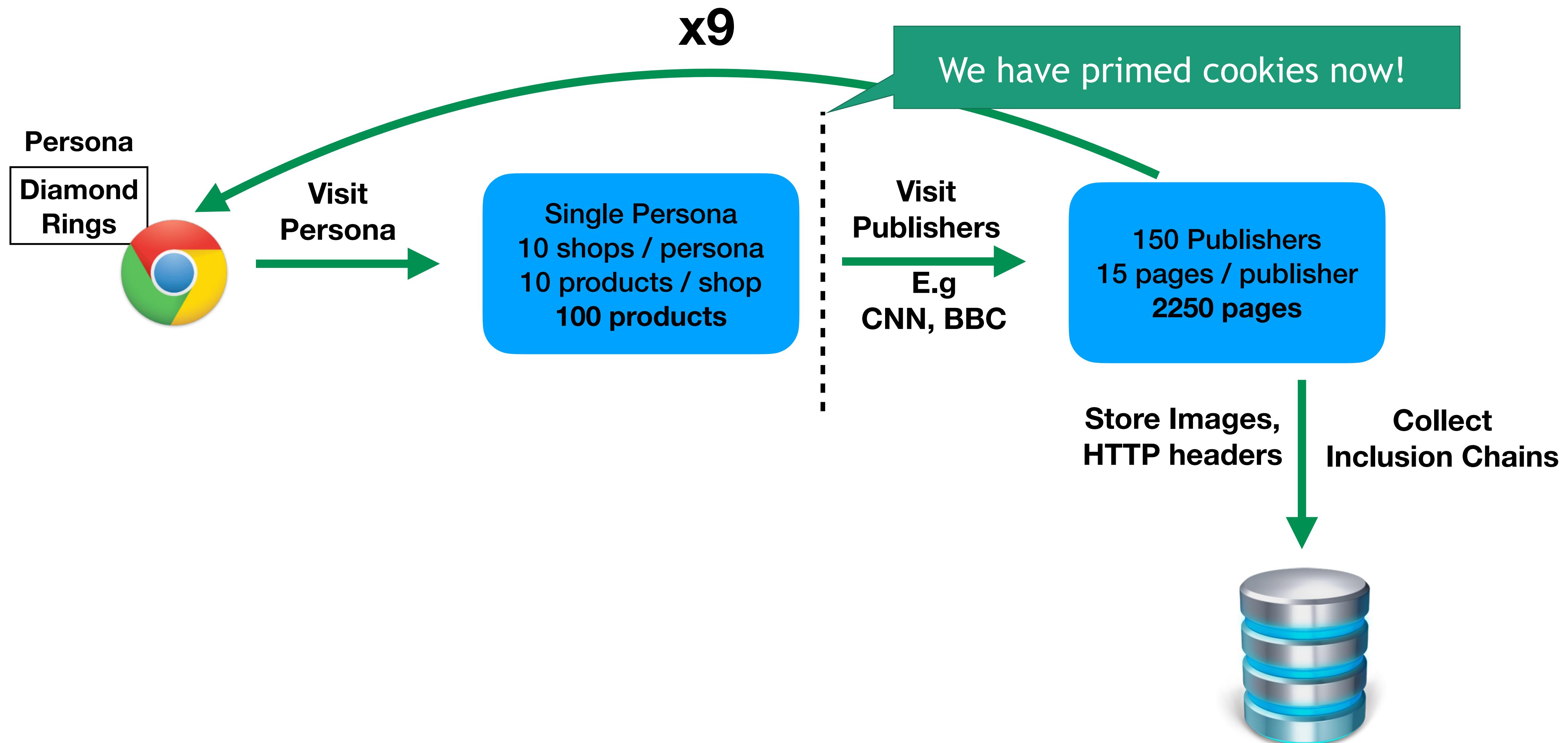
Experiment Setup & Data Collection



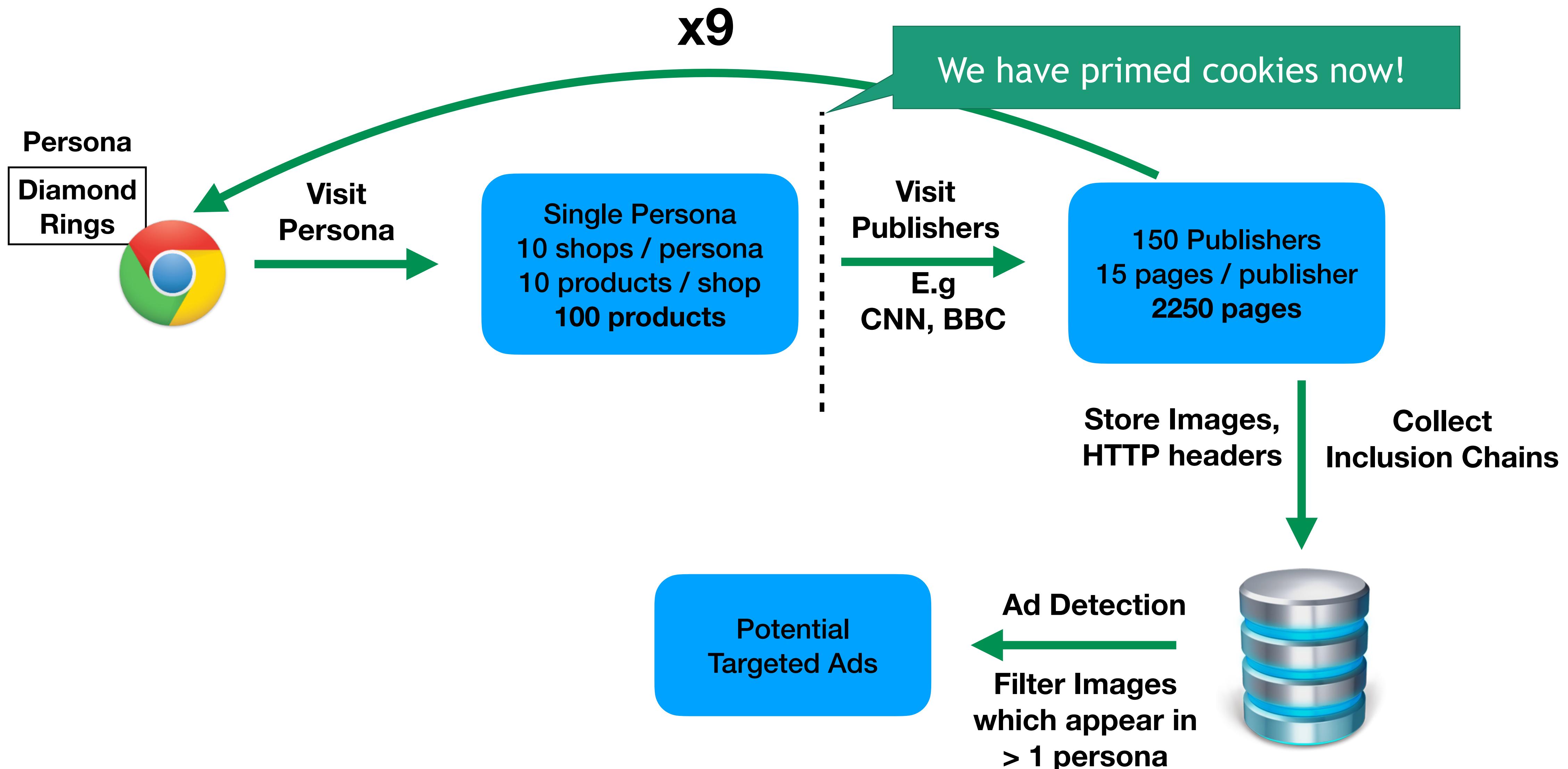
Experiment Setup & Data Collection



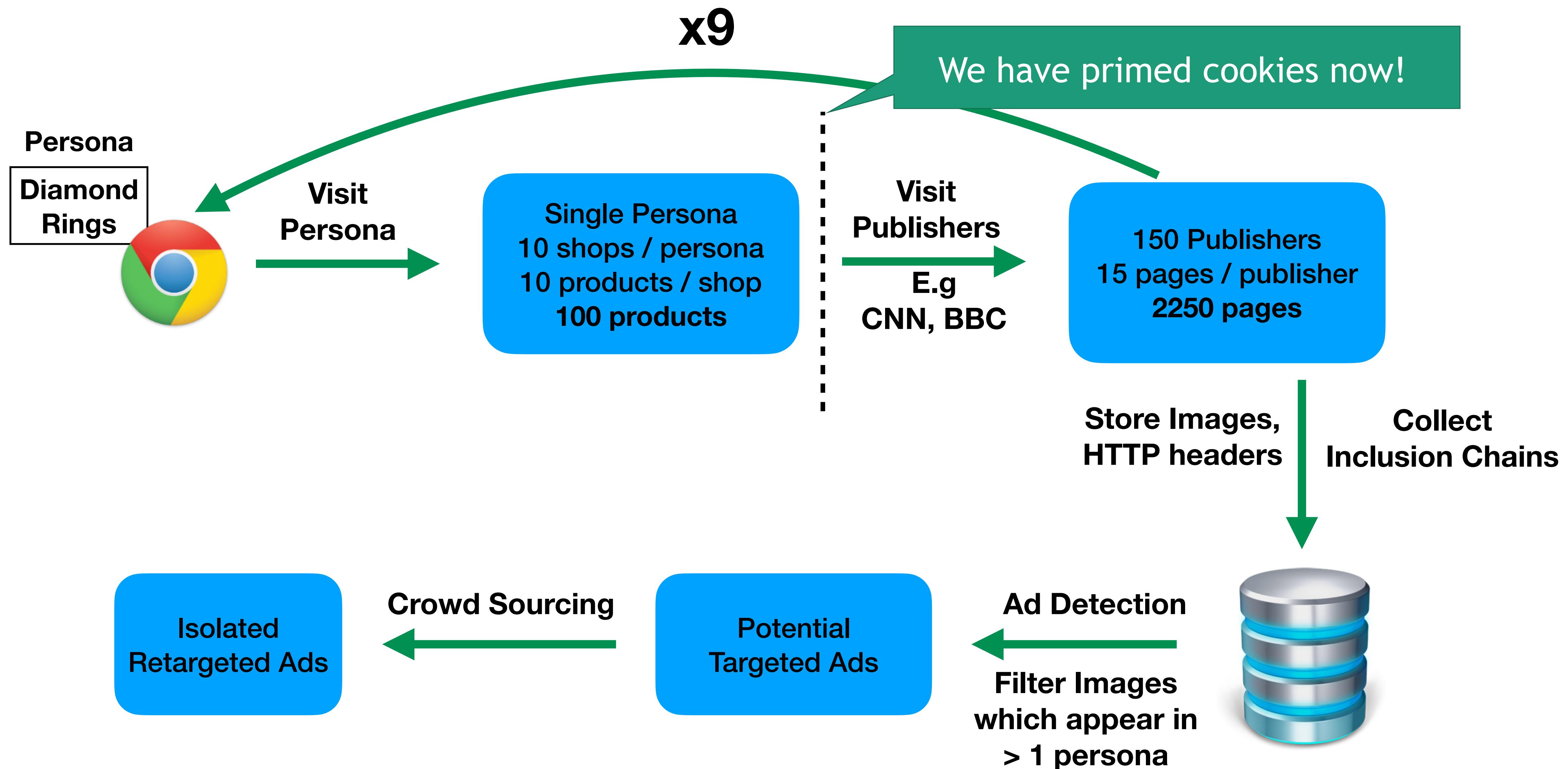
Experiment Setup & Data Collection



Experiment Setup & Data Collection

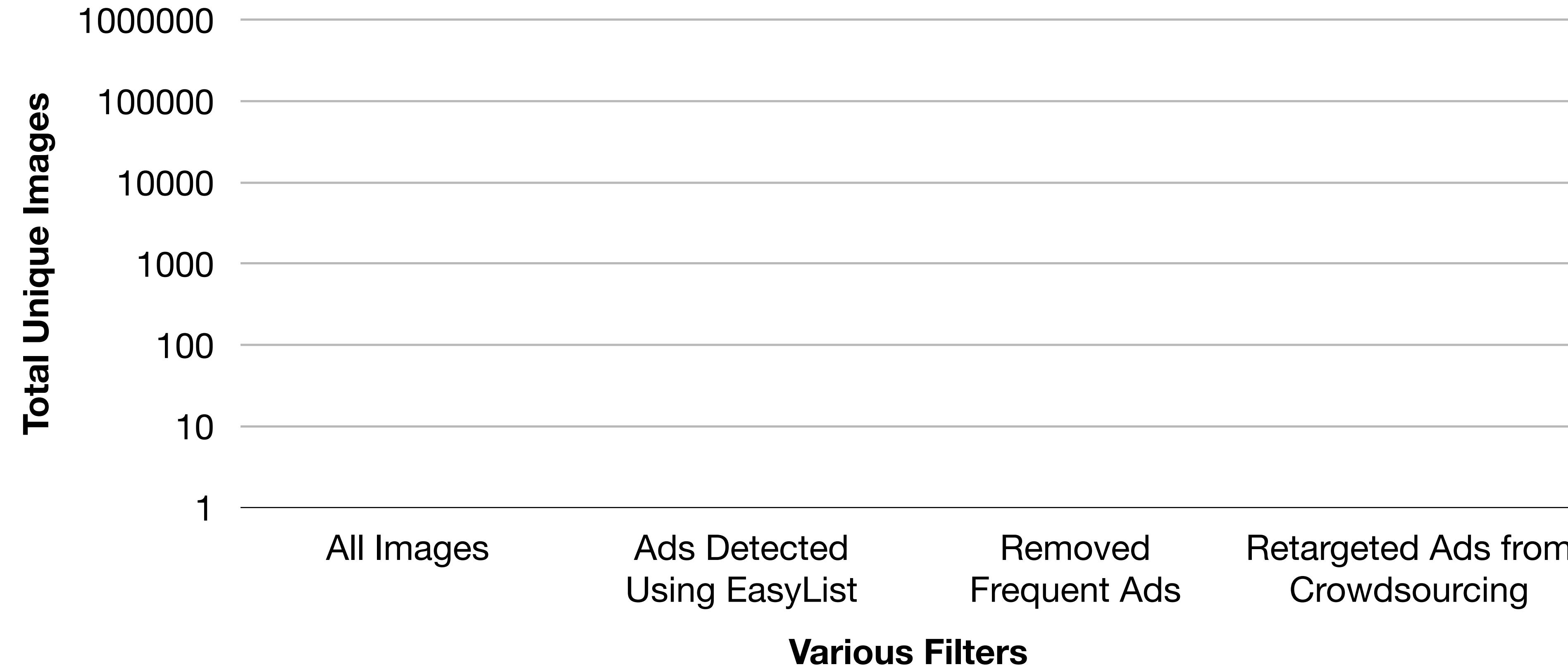


Experiment Setup & Data Collection

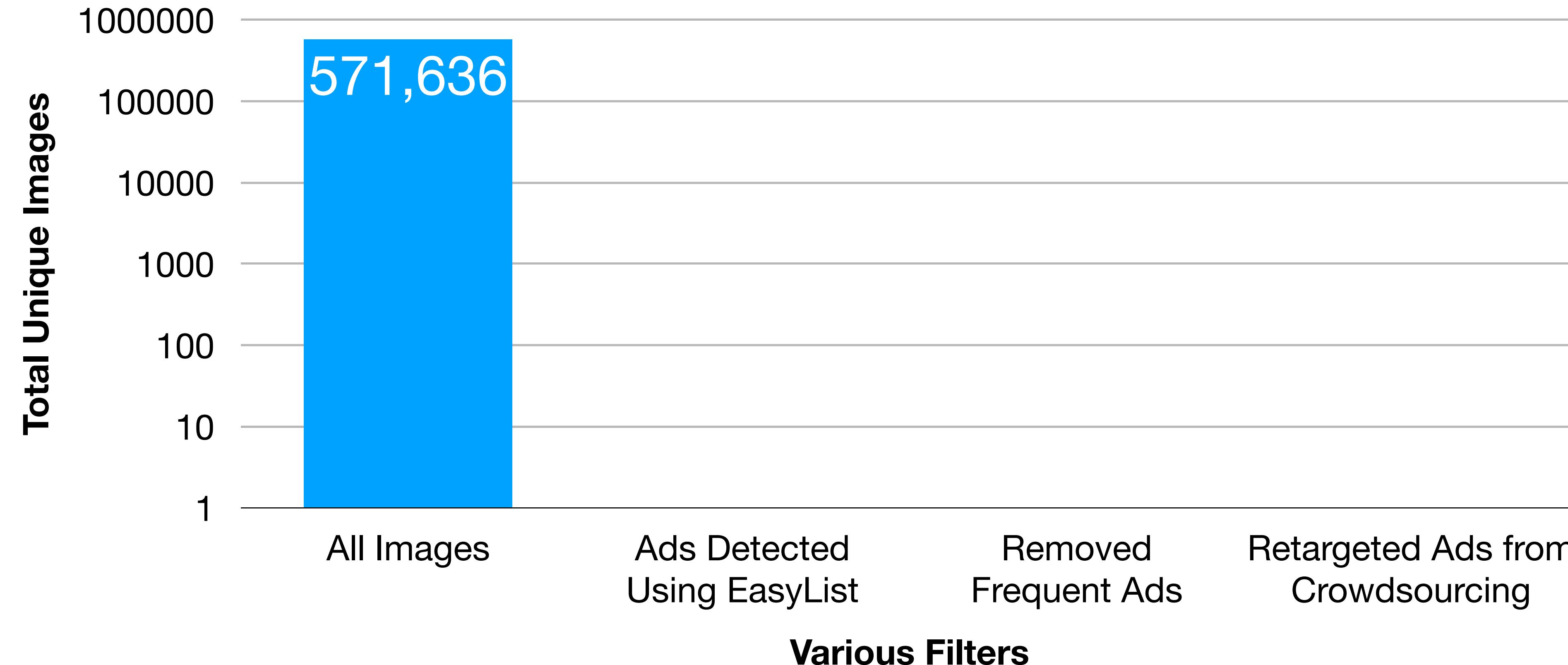


Overall Numbers

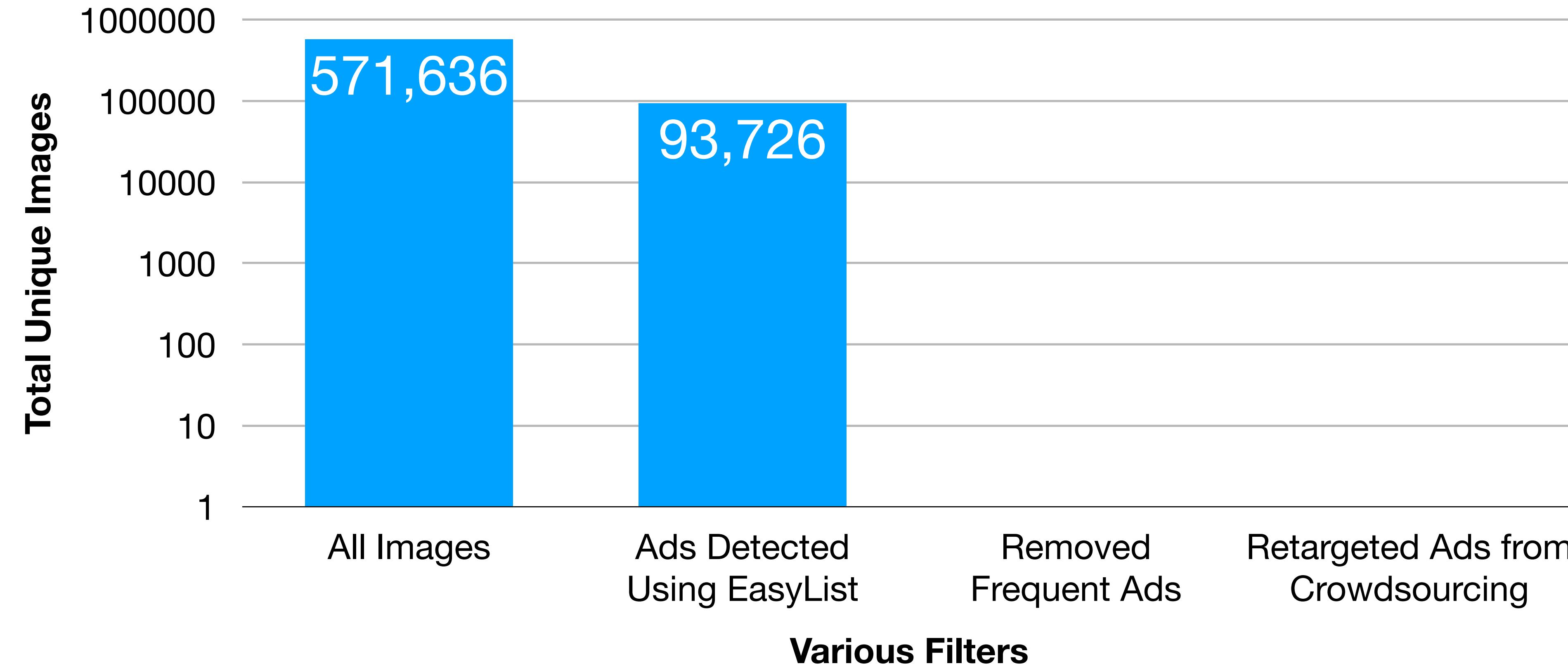
Overall Numbers



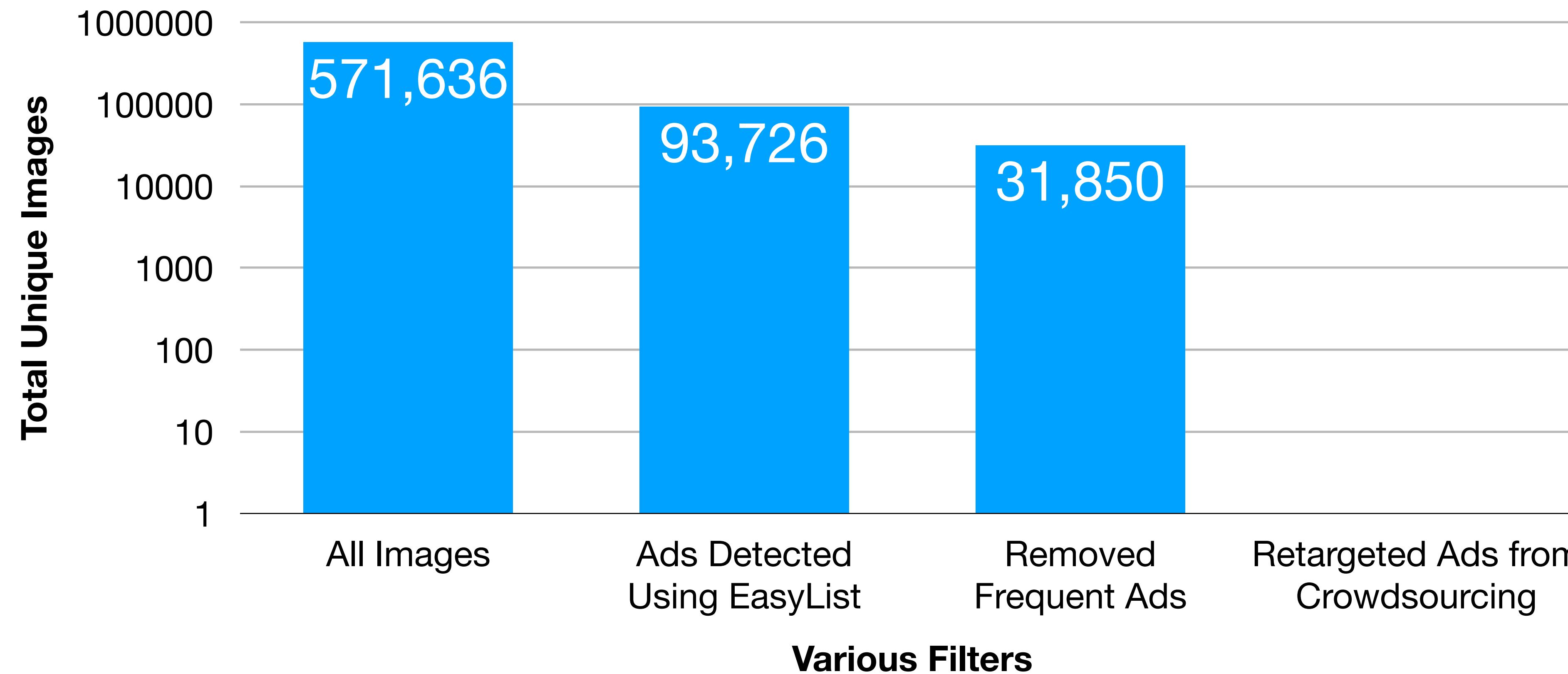
Overall Numbers



Overall Numbers

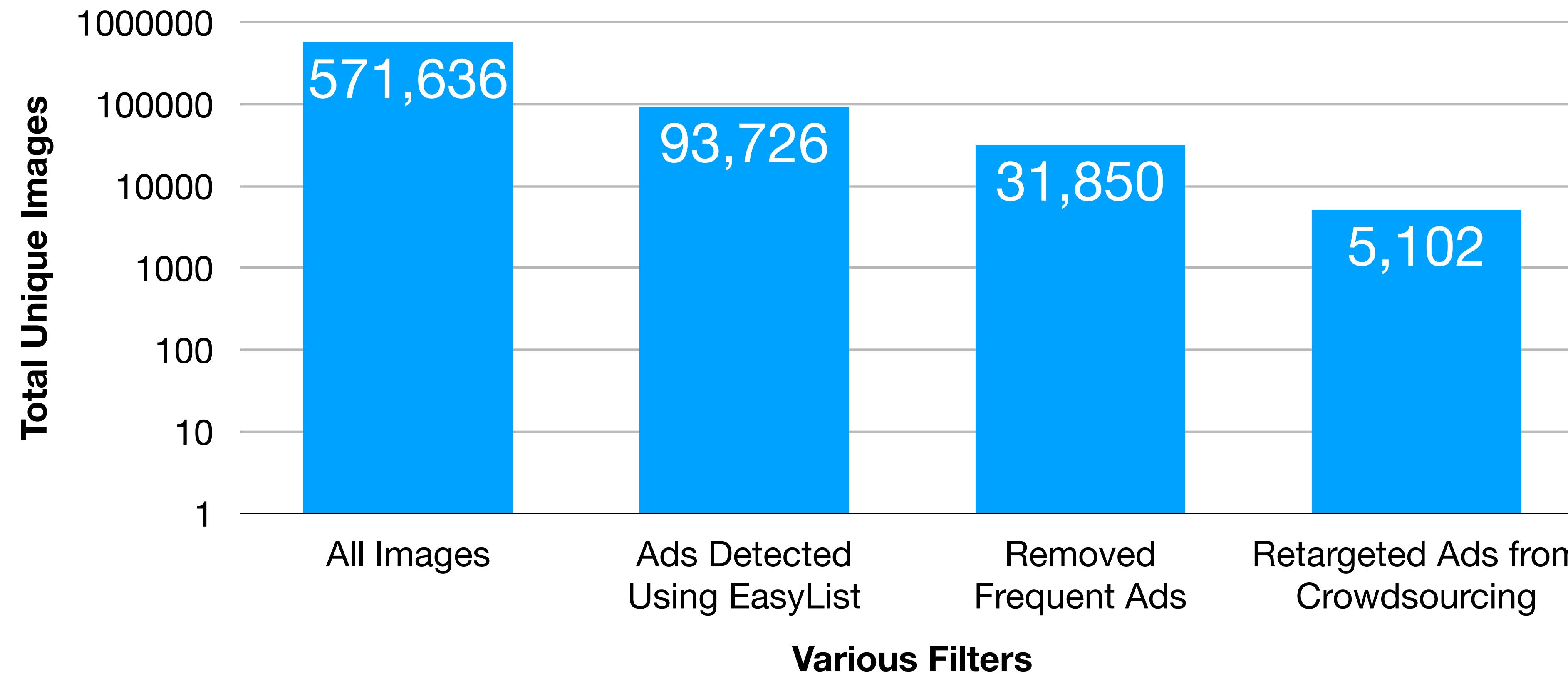


Overall Numbers



- Personas visited non-overlapping retailers
 - By definition, retargets should only be shown to a single persona
- Divided ~32K Images among 1,142 Tasks on Amazon Mechanical Turk (AMT)
 - Each task asked the worker to label 30 ad images

Overall Numbers



- Personas visited non-overlapping retailers
 - By definition, retargets should only be shown to a single persona
- Divided ~32K Images among 1,142 Tasks on Amazon Mechanical Turk (AMT)
 - Each task asked the worker to label 30 ad images

Example AMT Task

Your response	Image
<p>Select the appropriate category for the image.</p> <p><input checked="" type="radio"/> shopping <input type="radio"/> jewelry <input type="radio"/> diamonds Rings, necklace, etc.</p> <p><input type="radio"/> None of the above</p>	<p>The image shows a promotional advertisement for Adiamor. At the top, it says "You Design It • We Build It". Below this are two diamond rings. Underneath the rings, the text reads "Free Shipping • 30-Day Returns". At the bottom left is the Adiamor logo with the tagline "DIAMONDS & FINE JEWELRY", and at the bottom right is a "Shop Now" button.</p>

Example AMT Task

Your response	Image
<p>Select the appropriate category for the image.</p> <p><input checked="" type="radio"/> shopping <input type="radio"/> jewelry <input type="radio"/> diamonds Rings, necklace, etc.</p> <p><input type="radio"/> None of the above</p>	 <p>The image shows an advertisement for Adiamor Diamonds & Fine Jewelry. It features two diamond rings: one with a large central diamond and a smaller diamond band, and another with a large central diamond surrounded by smaller diamonds. The text "You Design It • We Build It" is at the top, followed by "Free Shipping • 30-Day Returns". The Adiamor logo is at the bottom left, and a "Shop Now" button is on the right.</p>
<p>Save and Continue</p>	

- We know in which persona we got a particular image.

Example AMT Task

Your response	Image
<p>Select the appropriate category for the image.</p> <p><input checked="" type="radio"/> shopping <input type="radio"/> jewelry <input type="radio"/> diamonds Rings, necklace, etc.</p> <p><input type="radio"/> None of the above</p> <p>Does this image say it came from one the following websites?</p> <p><input type="radio"/> adiamor.com <input type="radio"/> bluenile.com <input type="radio"/> gillets.com.au <input type="radio"/> lumeradiamonds.com <input type="radio"/> shaneco.com <input type="radio"/> szul.com <input type="radio"/> totaram.com <input type="radio"/> washingtondiamond.com <input type="radio"/> whiteflash.com <input type="radio"/> No</p> <p><input type="button" value="Save and Continue"/></p>	

- We know in which persona we got a particular image.
- We know the shops (websites) we visited per persona.

Example AMT Task

Your response	Image
<p>Select the appropriate category for the image.</p> <p><input checked="" type="radio"/> shopping <input type="radio"/> jewelry <input type="radio"/> diamonds Rings, necklace, etc.</p> <p><input type="radio"/> None of the above</p> <p>Does this image say it came from one the following websites?</p> <p><input checked="" type="radio"/> adiamor.com <input type="radio"/> bluenile.com <input type="radio"/> gillets.com.au <input type="radio"/> lumeradiamonds.com <input type="radio"/> shaneco.com <input type="radio"/> szul.com <input type="radio"/> totaram.com <input type="radio"/> washingtondiamond.com <input type="radio"/> whiteflash.com <input type="radio"/> No</p>	

- We know in which persona we got a particular image.
- We know the shops (websites) we visited per persona.

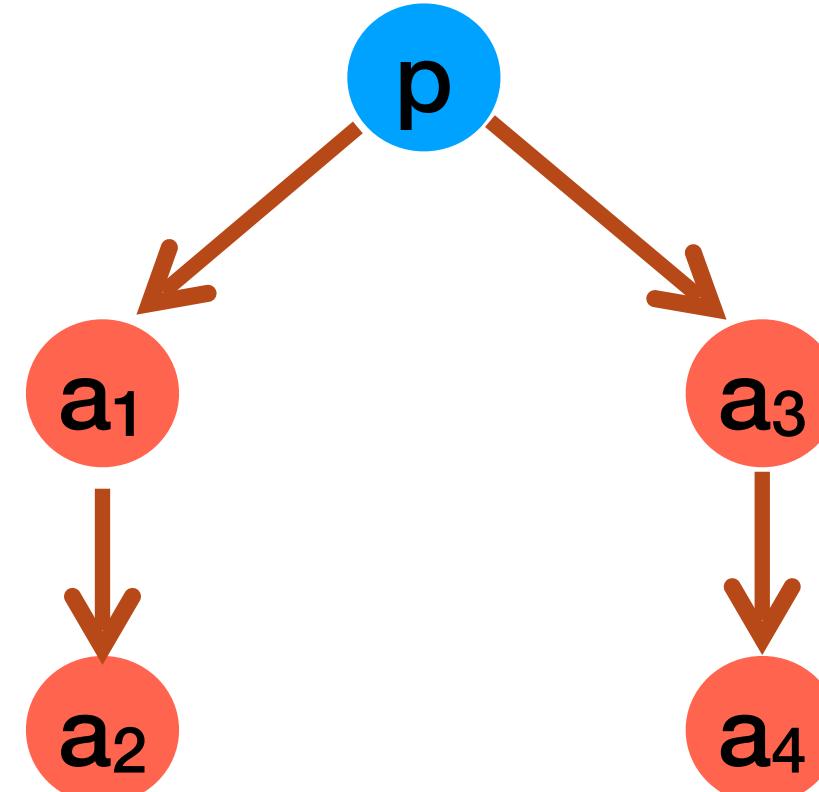
Inclusion Chains Associated With Retargeted Ads

DOM Tree for <http://p.com/index.html>

```
<html>
  <body>
    <script src="a1.com/cookie-match.js"></script>
    <!-- Tracking pixel inserted dynamically
        by cookie-match.js -->
    

    <iframe src="a3.com/banner.html">
      <script src="a4.com/ads.js"></script>
    </iframe>
  </body>
</html>
```

Inclusion Chains



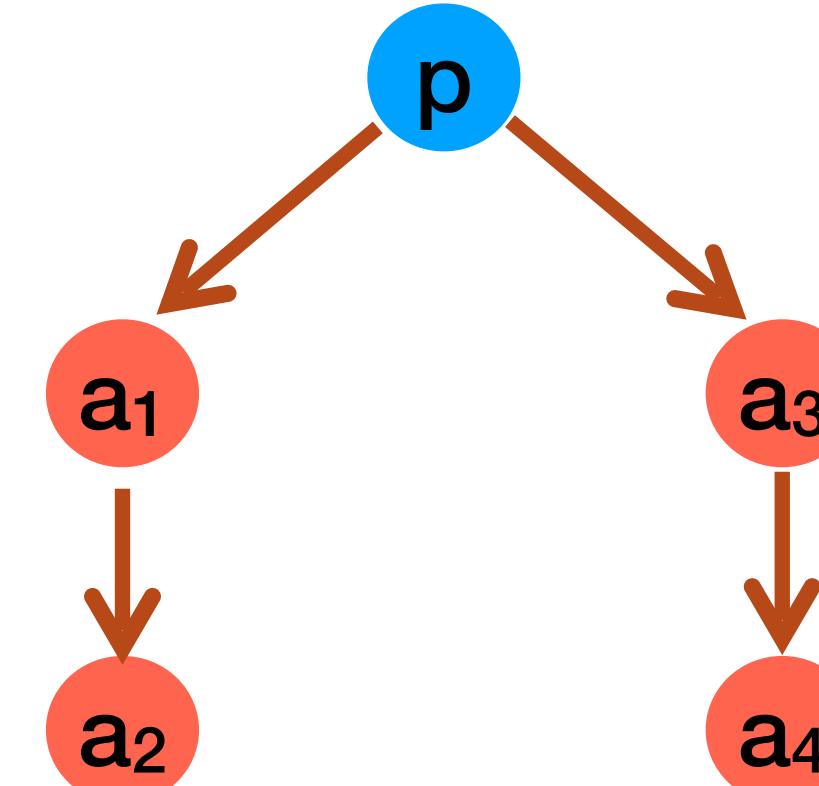
Inclusion Chains Associated With Retargeted Ads

DOM Tree for <http://p.com/index.html>

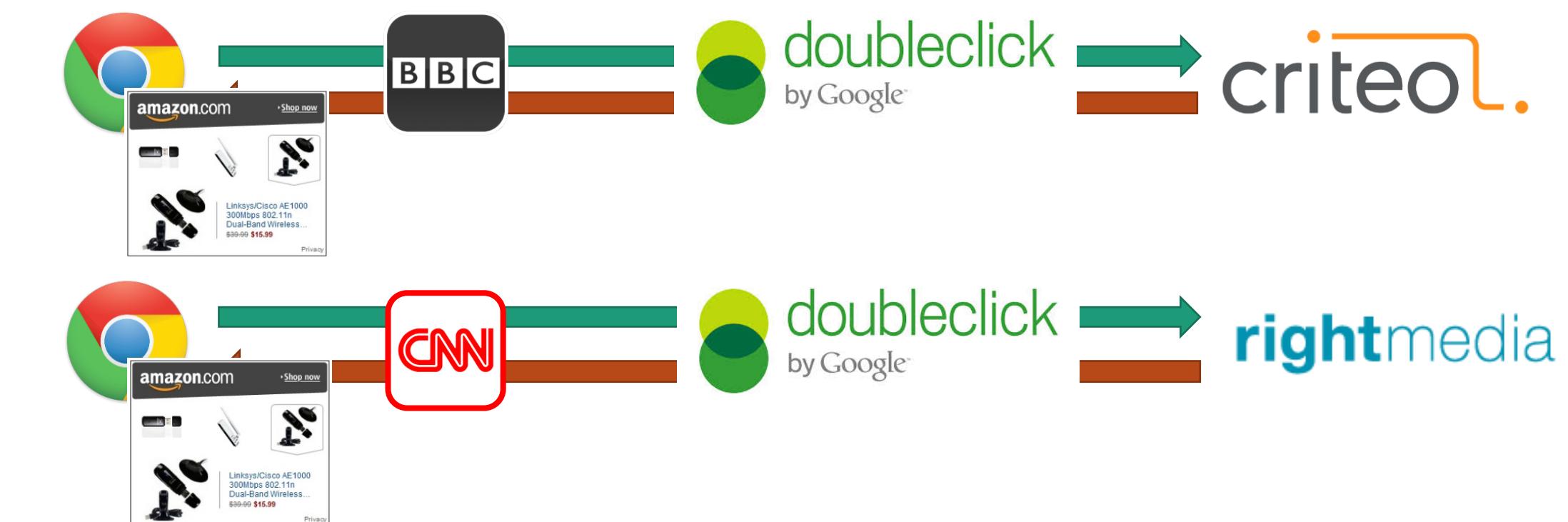
```
<html>
  <body>
    <script src="a1.com/cookie-match.js"></script>
    <!-- Tracking pixel inserted dynamically
        by cookie-match.js -->
    

    <iframe src="a3.com/banner.html">
      <script src="a4.com/ads.js"></script>
    </iframe>
  </body>
</html>
```

Inclusion Chains



- 5,102 unique retargeted ads
- 35,448 associated publisher-side chains
 - We observed some retargets multiple times



An Example of Publisher & Shopper Side Chains

An Example of Publisher & Shopper Side Chains

Example

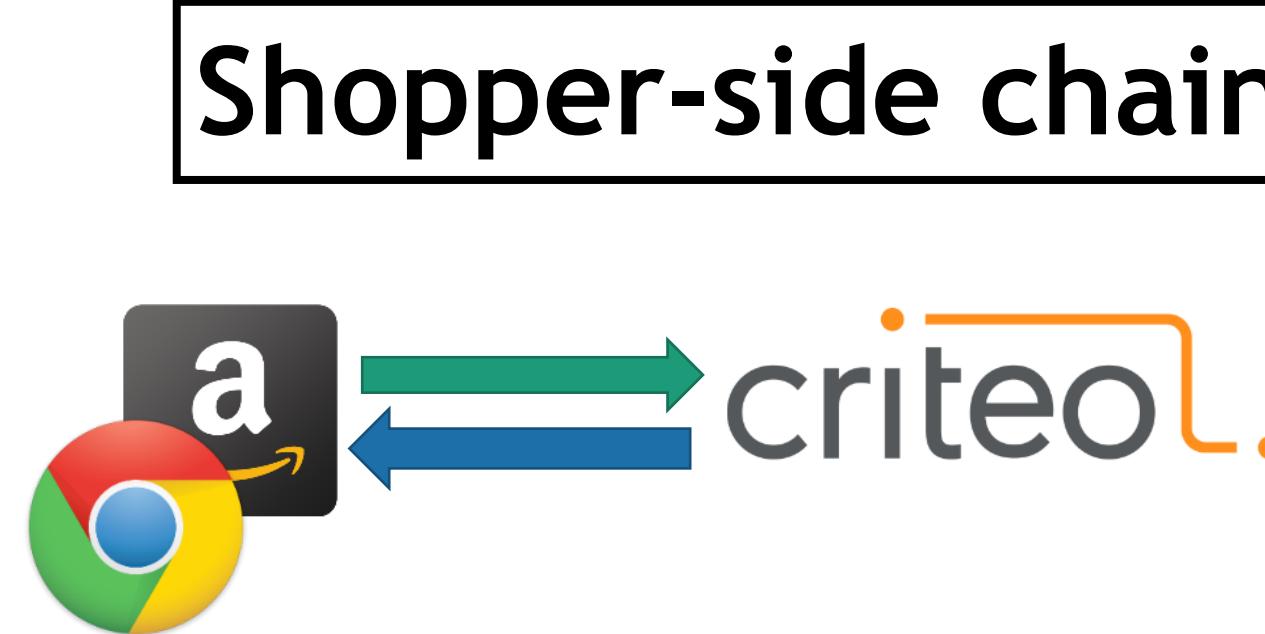
Let's see what caused this Ad to be shown to us?

Publisher-side chain



An Example of Publisher & Shopper Side Chains

Example



Let's see what caused this Ad to be shown to us?

Publisher-side chain

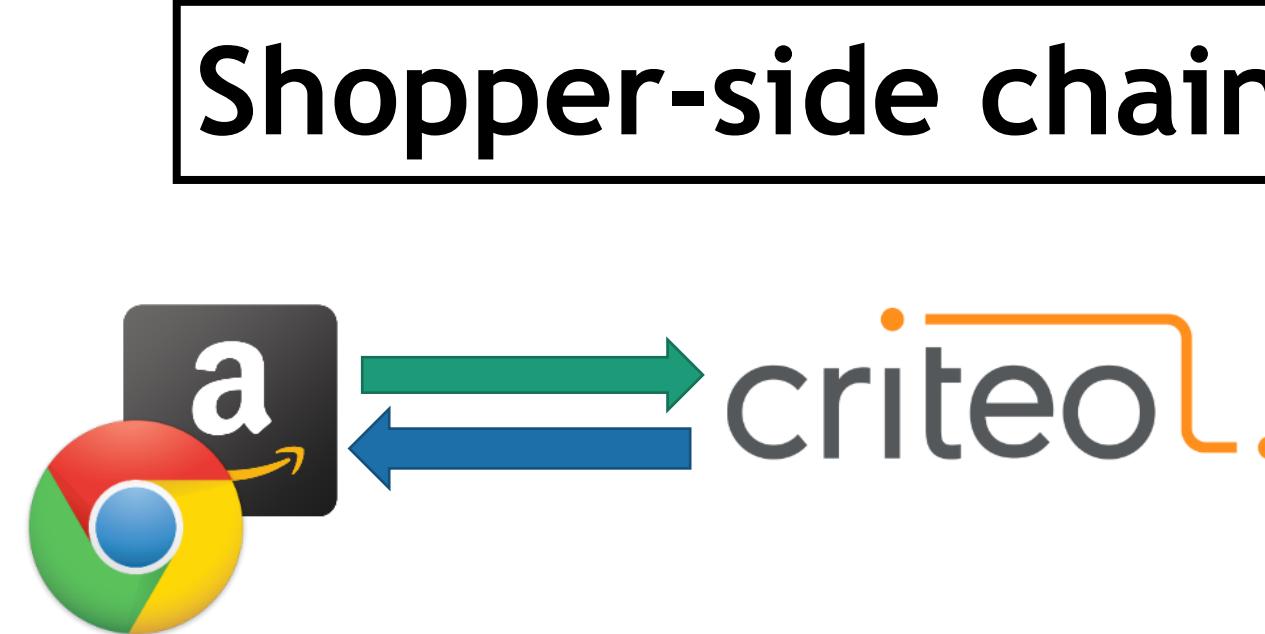


How does criteo know to serve ad on BBC ?

- In this case it is pretty trivial
- criteo observed the user at the (shop) a

An Example of Publisher & Shopper Side Chains

Example



Let's see what caused this Ad to be shown to us?

Publisher-side chain



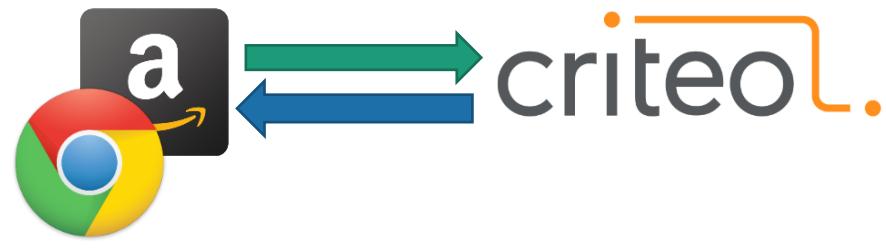
How does criteo know to serve ad on BBC ?

- In this case it is pretty trivial
- criteo observed the user at the (shop) a

Challenge: Can we classify all such publisher-side chains formally?

A Pattern for Chain Classification

Shopper-side chain



Publisher-side chain



- d Demand Side Platform (DSP)
- e Exchange (Running Auction)

A Pattern for Chain Classification

Shopper-side chain



Publisher-side chain



- d Demand Side Platform (DSP)
- e Exchange (Running Auction)

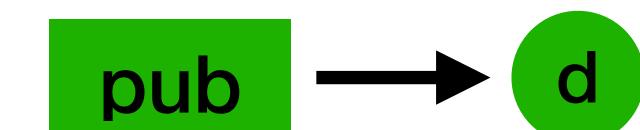
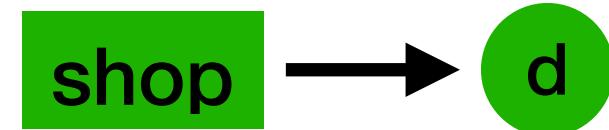
A Pattern for Chain Classification

-  **Demand Side Platform (DSP)**
-  **Exchange (Running Auction)**

Shopper-side chain



Publisher-side chain



A Pattern for Chain Classification

-  d Demand Side Platform (DSP)
-  e Exchange (Running Auction)

Shopper-side chain



Publisher-side chain



shop → d

pub → d

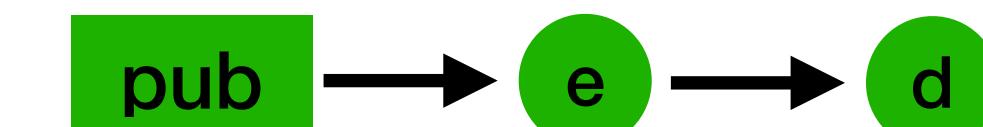
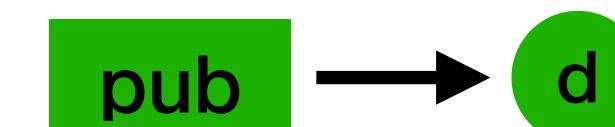
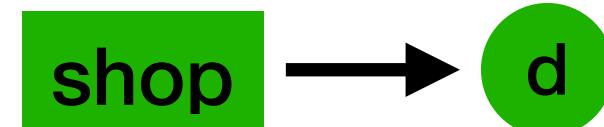
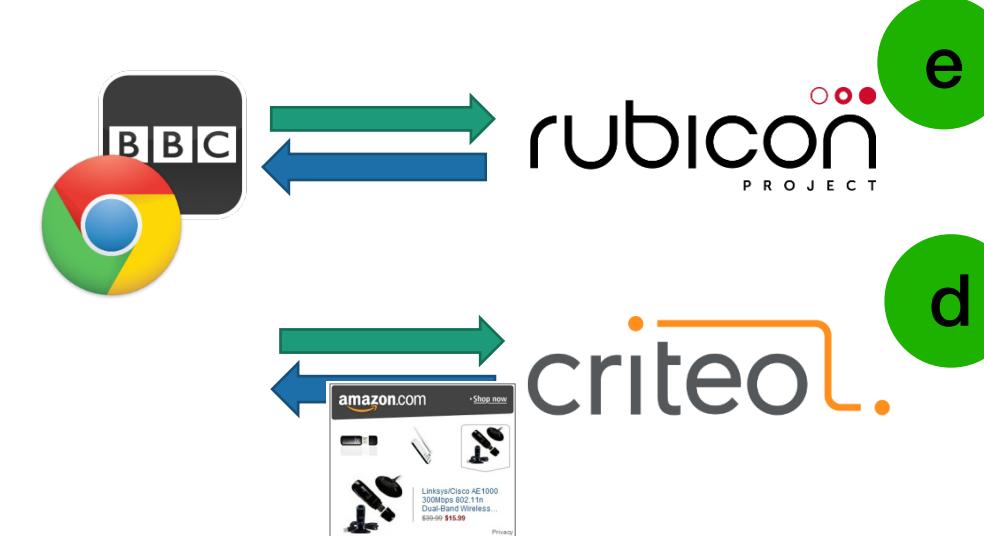
A Pattern for Chain Classification

-  Demand Side Platform (DSP)
-  Exchange (Running Auction)

Shopper-side chain



Publisher-side chain



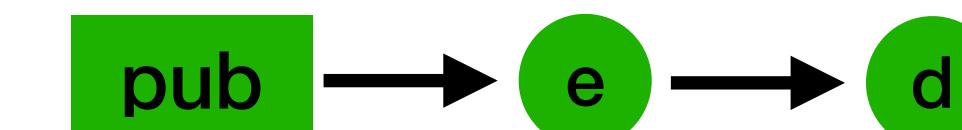
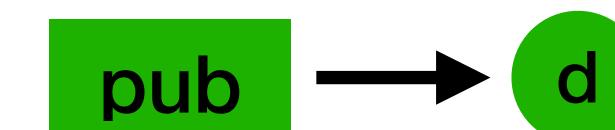
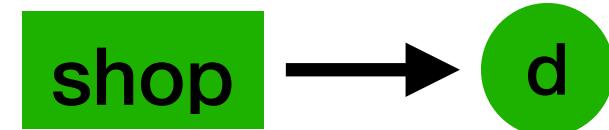
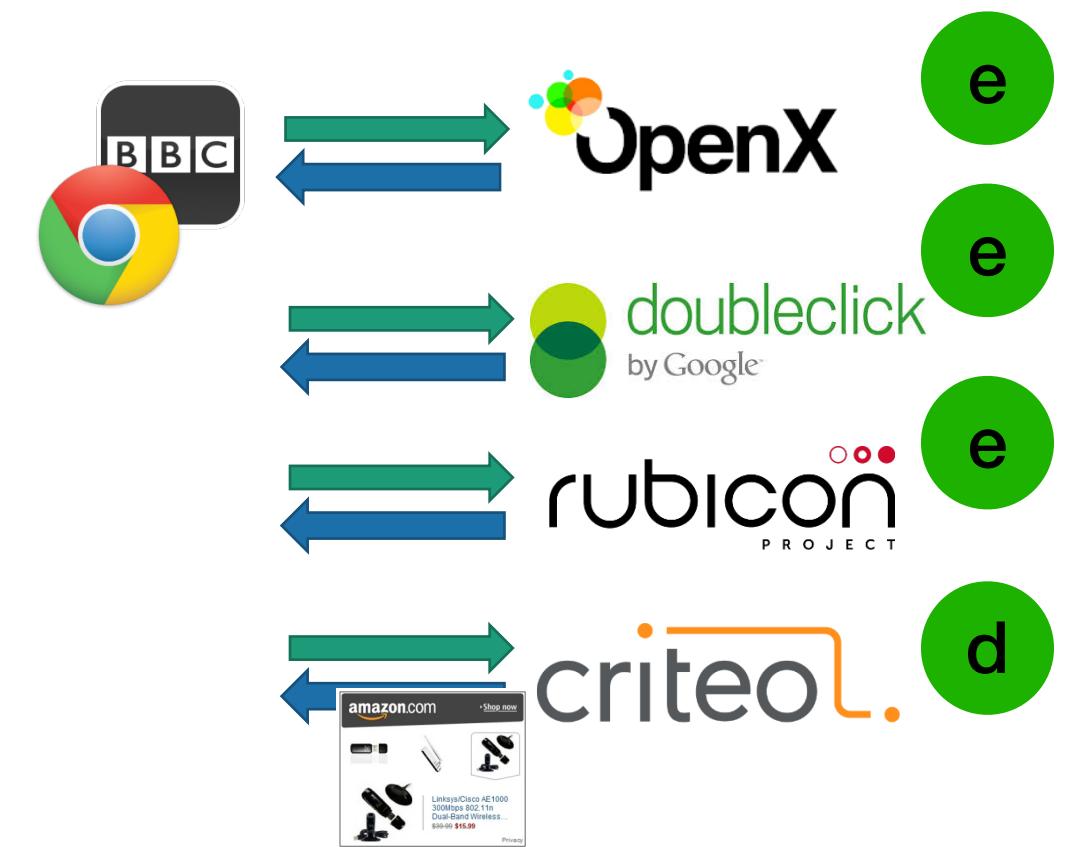
A Pattern for Chain Classification

-  Demand Side Platform (DSP)
-  Exchange (Running Auction)

Shopper-side chain



Publisher-side chain



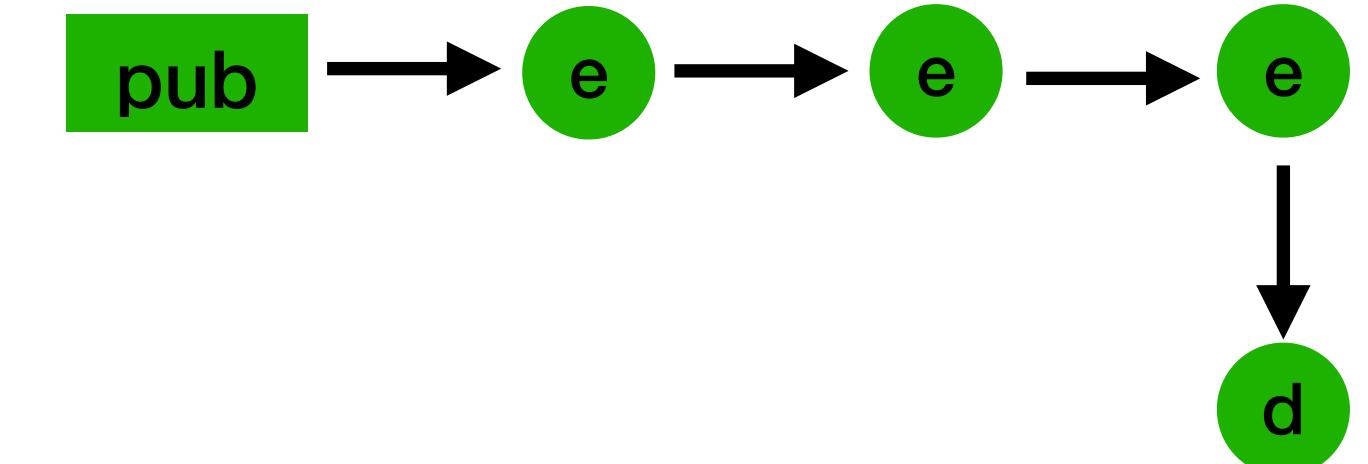
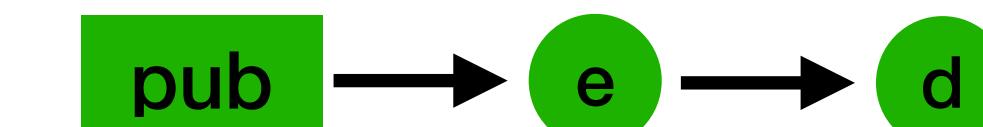
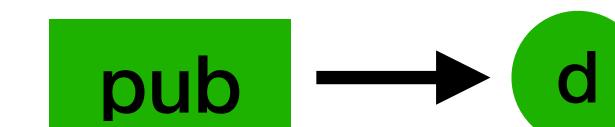
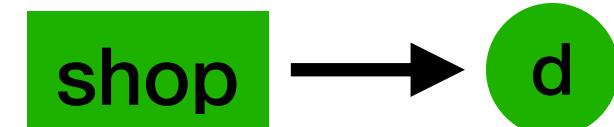
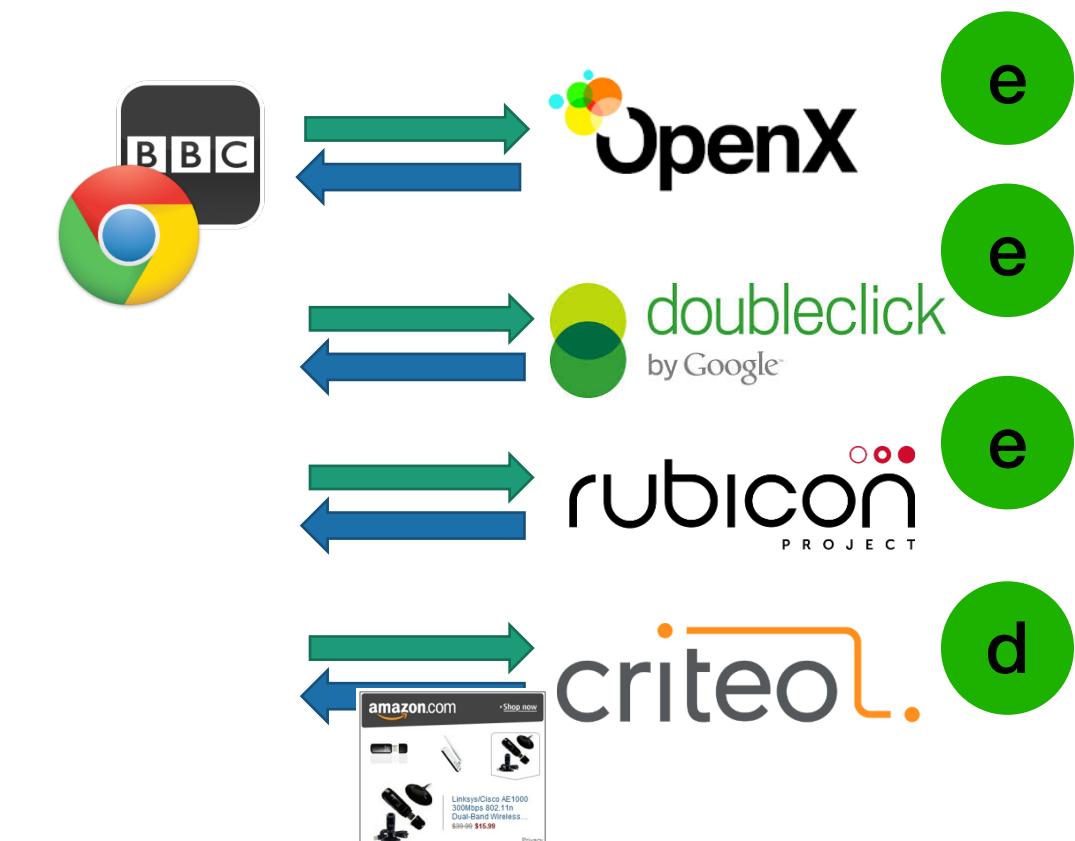
A Pattern for Chain Classification

-  Demand Side Platform (DSP)
-  Exchange (Running Auction)

Shopper-side chain



Publisher-side chain



A Pattern for Chain Classification

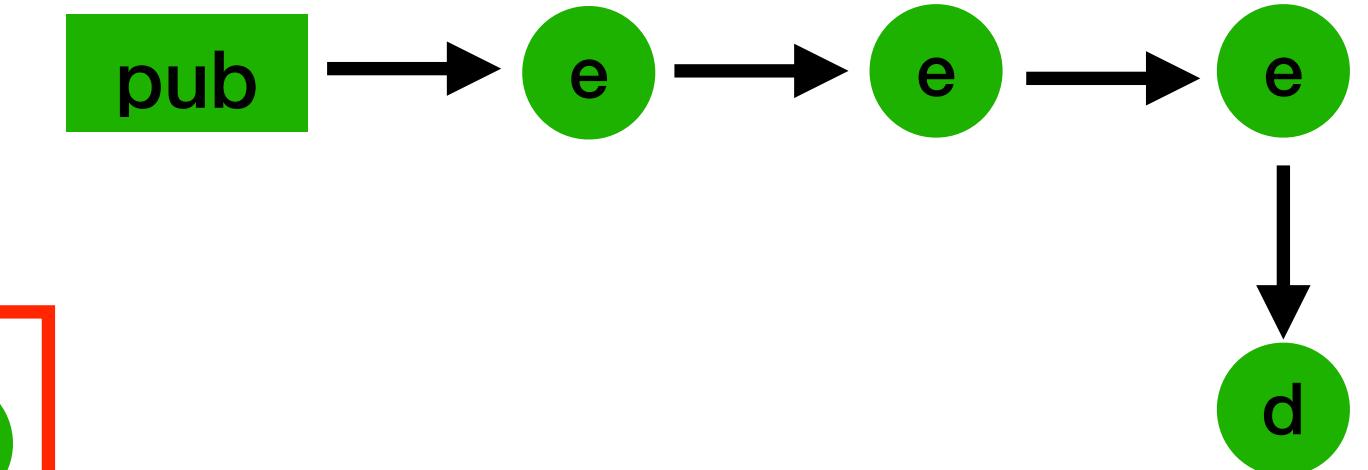
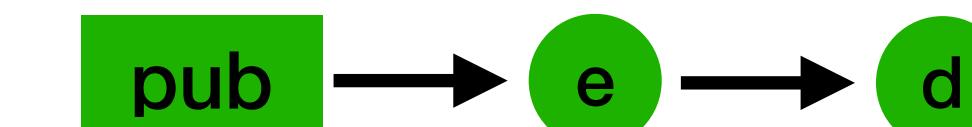
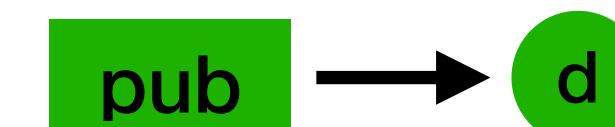
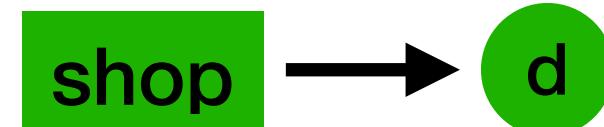
Shopper-side chain



Publisher-side chain



-  Demand Side Platform (DSP)
-  Exchange (Running Auction)

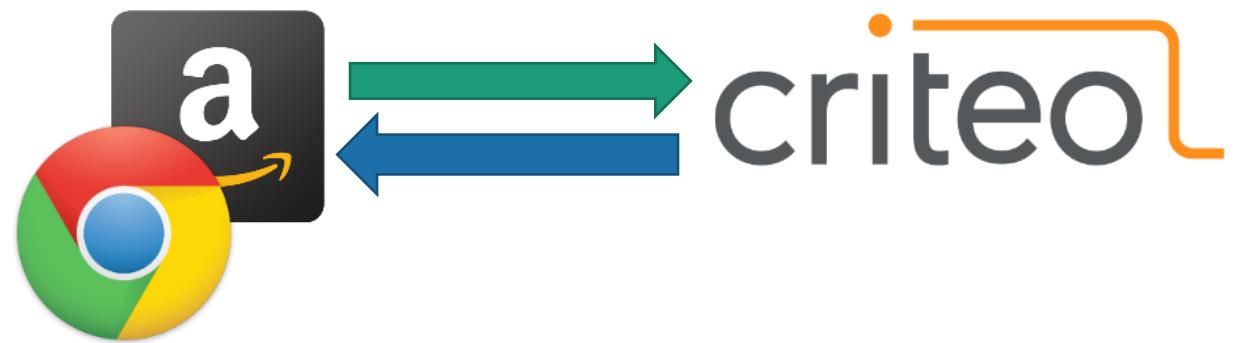


Regular Expression Like Rules

Direct (Trivial) Matching

Rule

Example



Shopper-side chain

Publisher-side chain

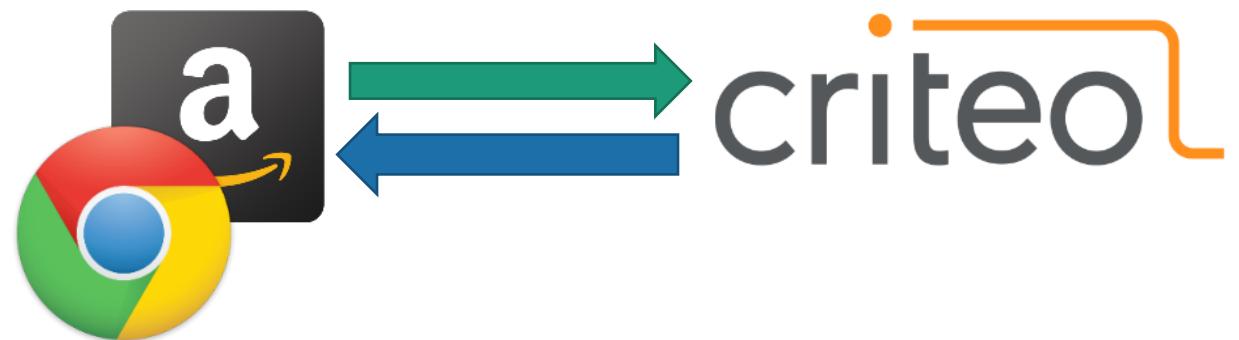


- d Demand Side Platform (DSP)
- e Exchange (Running Auction)

Direct (Trivial) Matching

Rule

Example



Shopper-side chain

Publisher-side chain



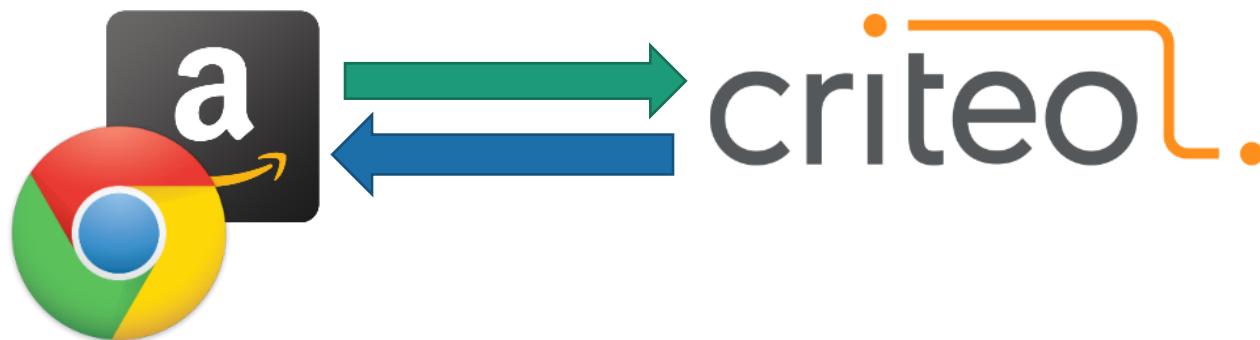
d is the DSP
that serves the
retarget

Direct (Trivial) Matching

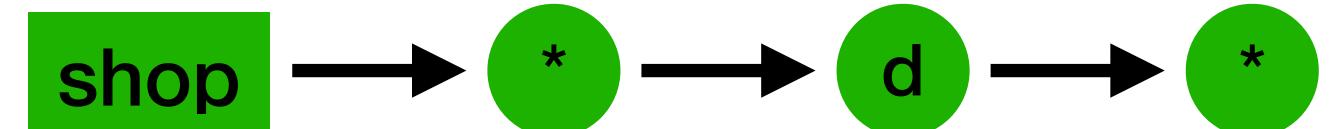


Example

Shopper-side chain



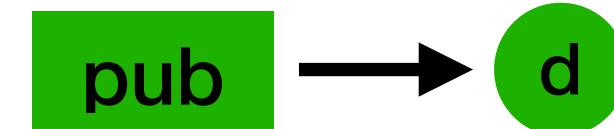
Rule



DSP d must observe us on the shopper side

... but other trackers may also appear

Publisher-side chain

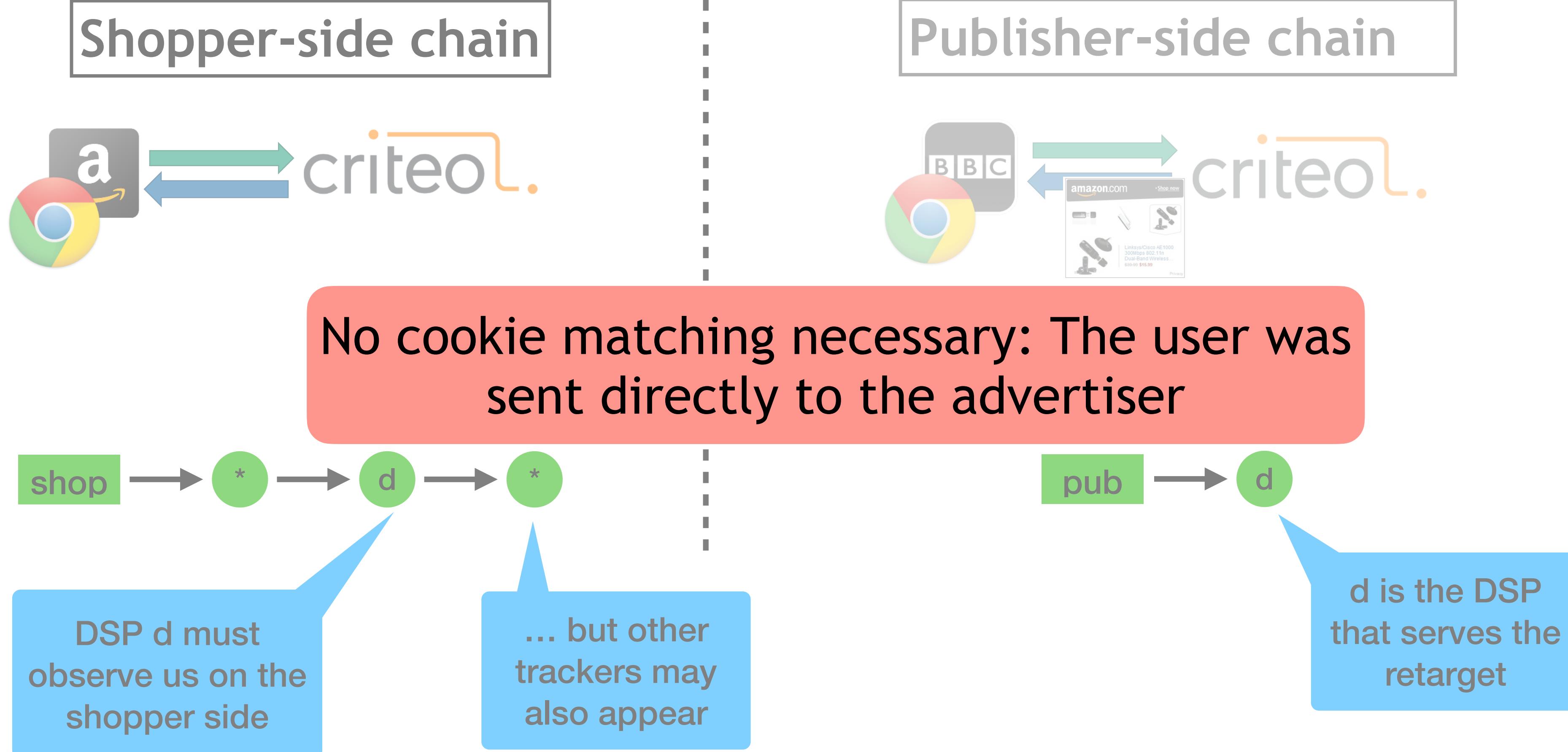


d is the DSP that serves the retarget

Direct (Trivial) Matching

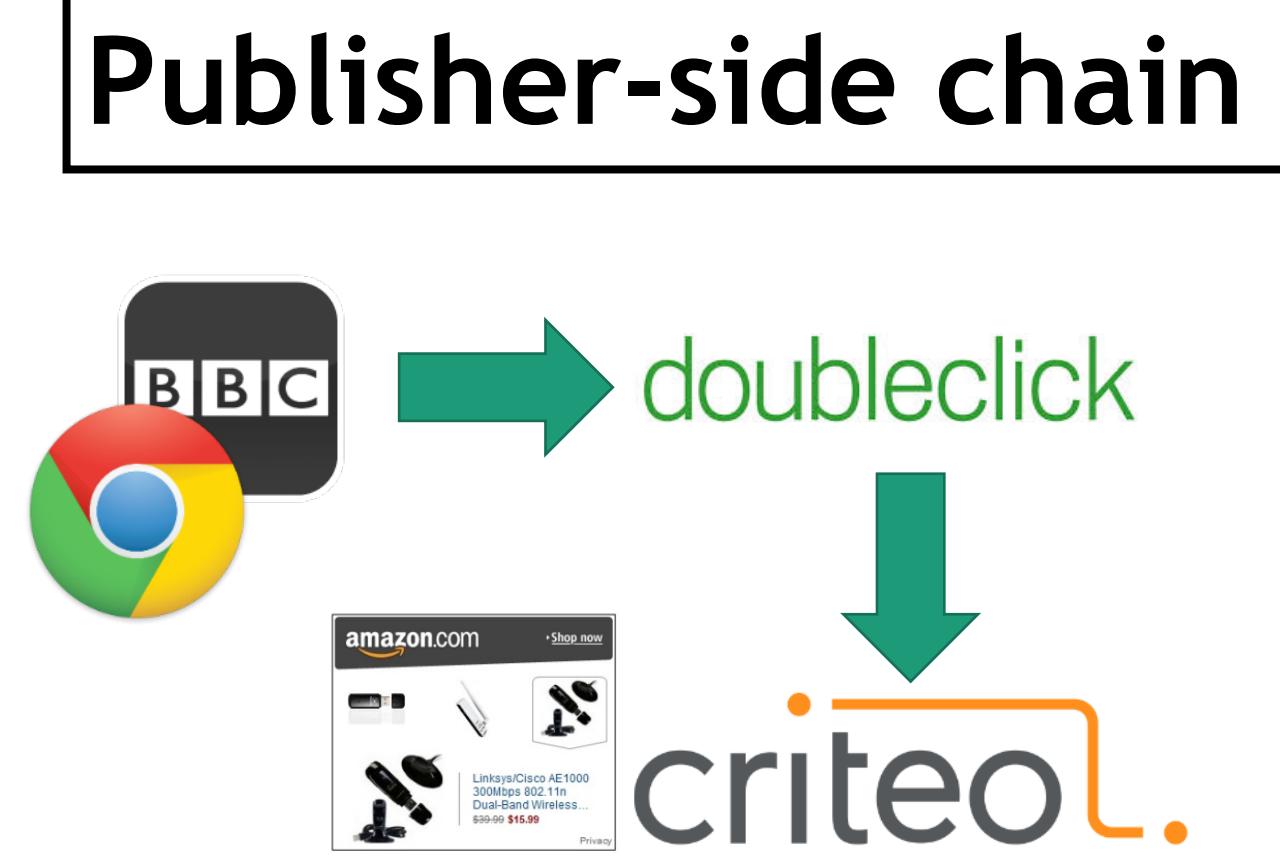
- d Demand Side Platform (DSP)
- e Exchange (Running Auction)

Example



Cookie Matching

Example Rule



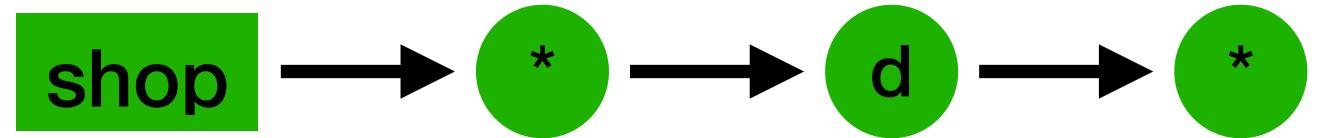
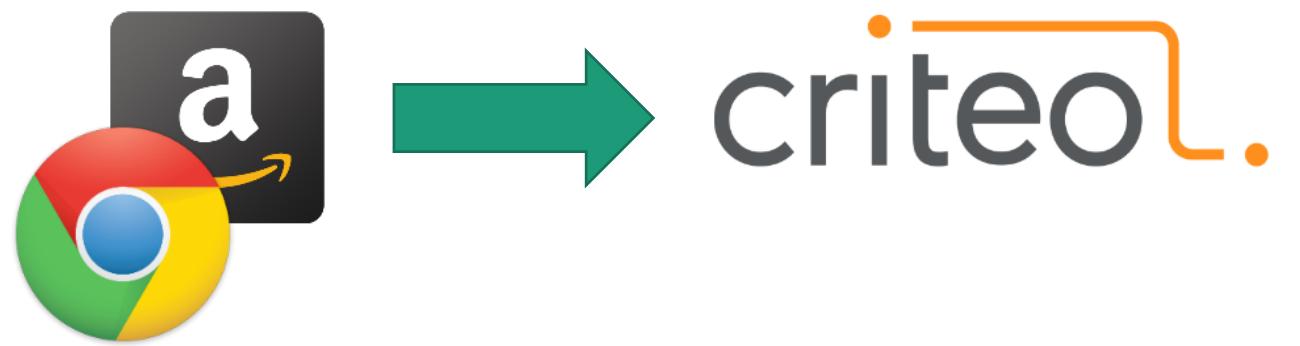
- d Demand Side Platform (DSP)
- e Exchange (Running Auction)

Cookie Matching

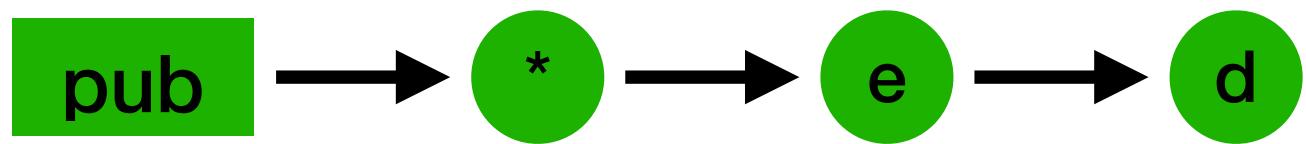
Example

Rule

Shopper-side chain



Publisher-side chain



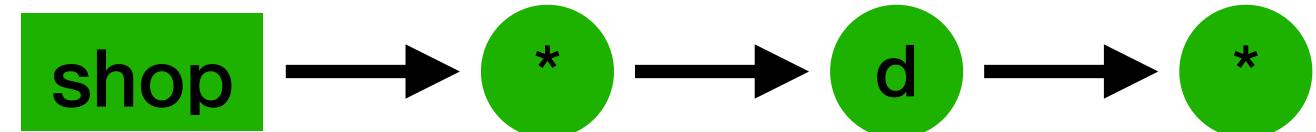
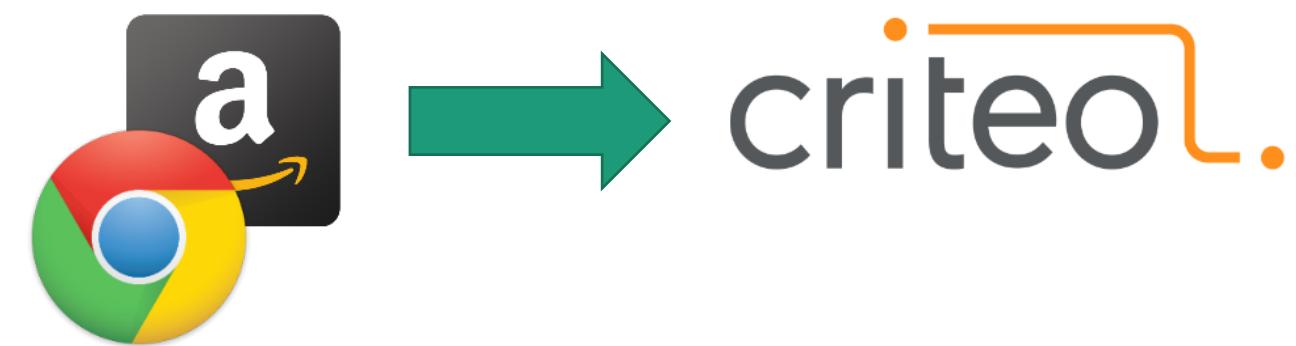
-  Demand Side Platform (DSP)
-  Exchange (Running Auction)

Cookie Matching

Example

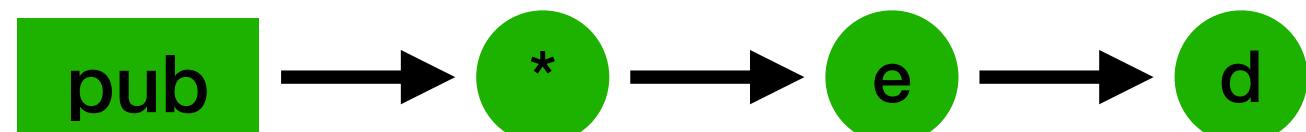
Rule

Shopper-side chain



DSP d must observe us on the shopper side

Publisher-side chain



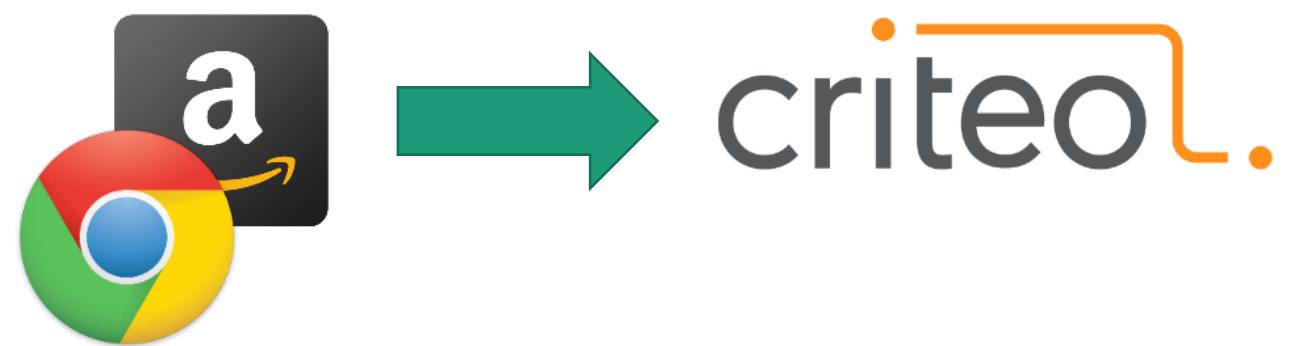
e precedes d, which implies an RTB auction

-  Demand Side Platform (DSP)
-  Exchange (Running Auction)

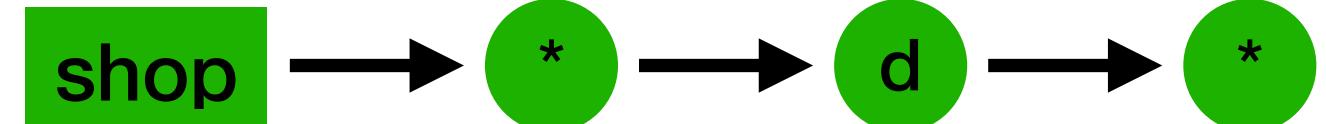
Cookie Matching

Example

Shopper-side chain

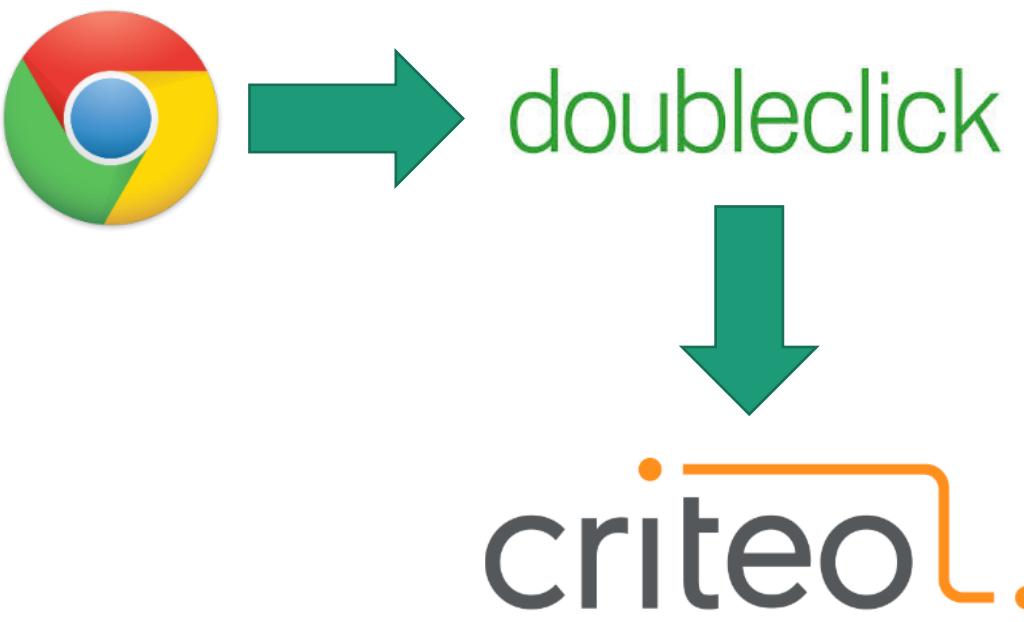


Rule



DSP d must observe us on the shopper side

Anywhere



Publisher-side chain



e precedes d, which implies an RTB auction

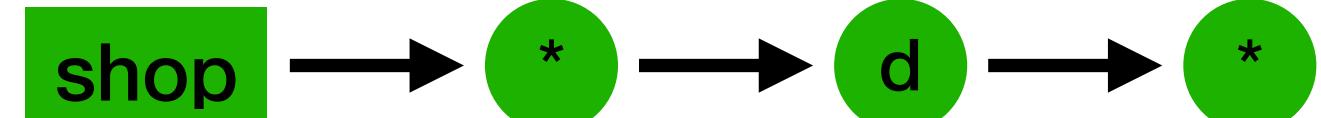
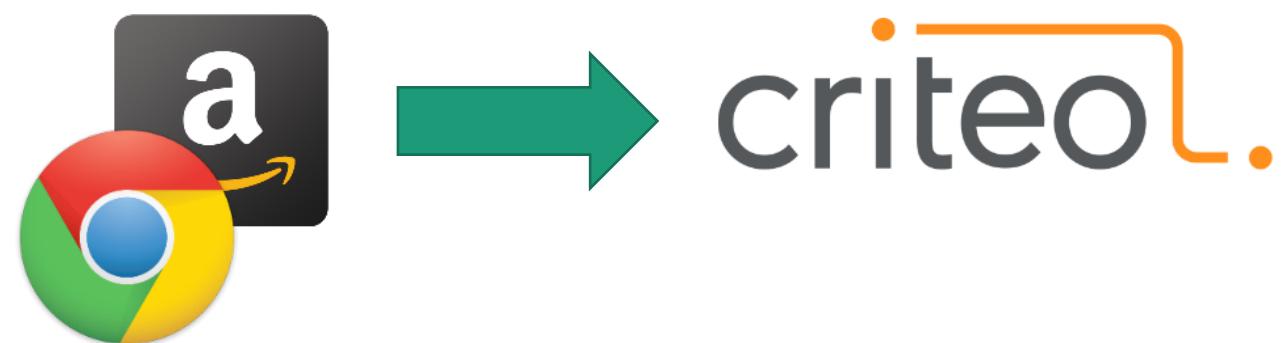
- d Demand Side Platform (DSP)
- e Exchange (Running Auction)

Cookie Matching

Example

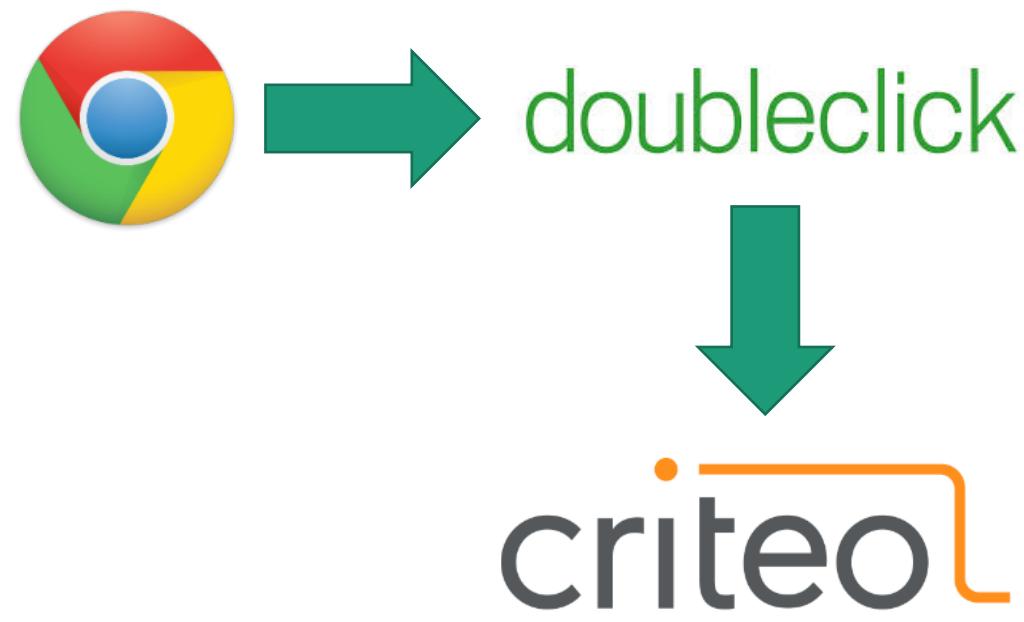
Rule

Shopper-side chain



DSP d must observe us on the shopper side

Anywhere



Publisher-side chain



e precedes d, which implies an RTB auction

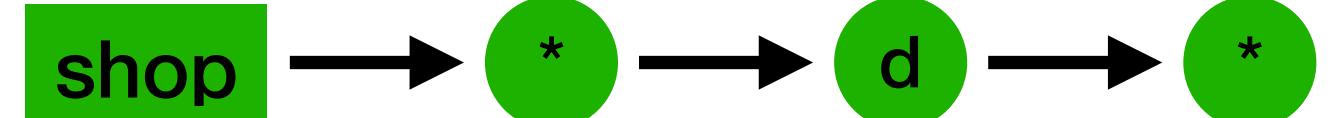
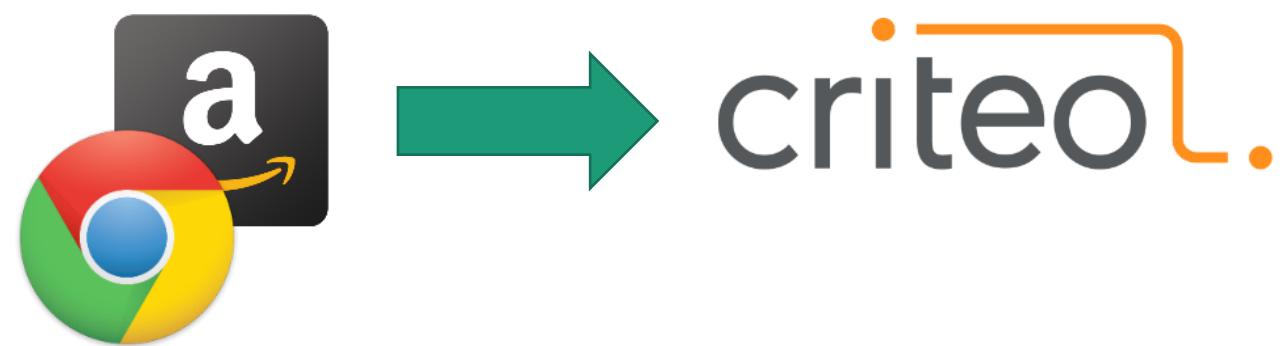
- d Demand Side Platform (DSP)
- e Exchange (Running Auction)

Cookie Matching

Example

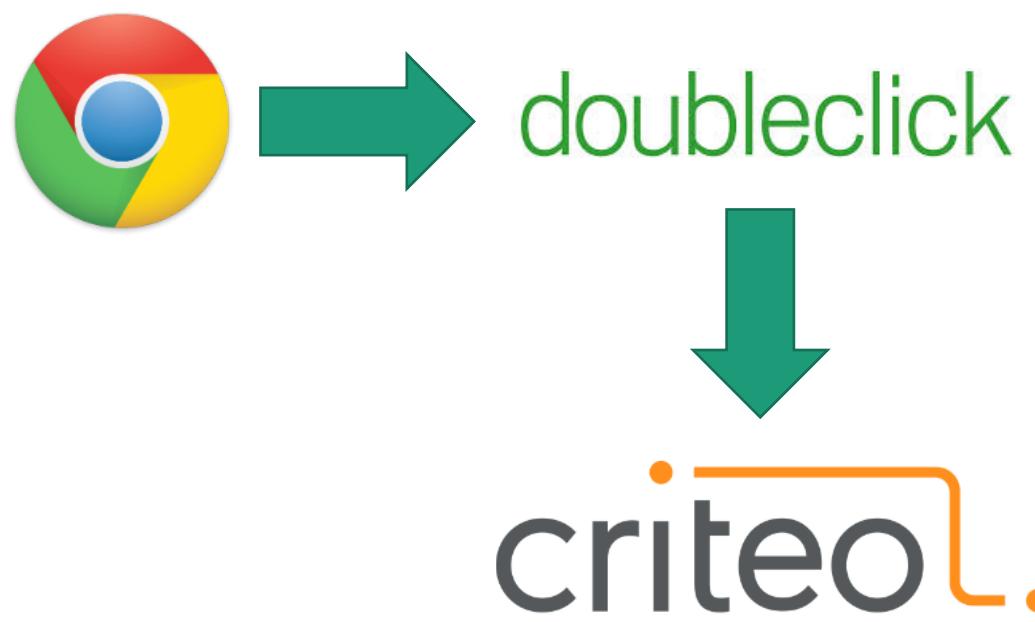
Rule

Shopper-side chain



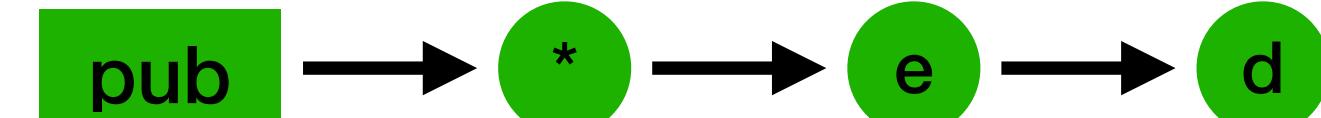
DSP d must observe us on the shopper side

Anywhere



Transition e → d is where cookie match happens

Publisher-side chain



e precedes d, which implies an RTB auction

-  Demand Side Platform (DSP)
-  Exchange (Running Auction)

Cookie Matching

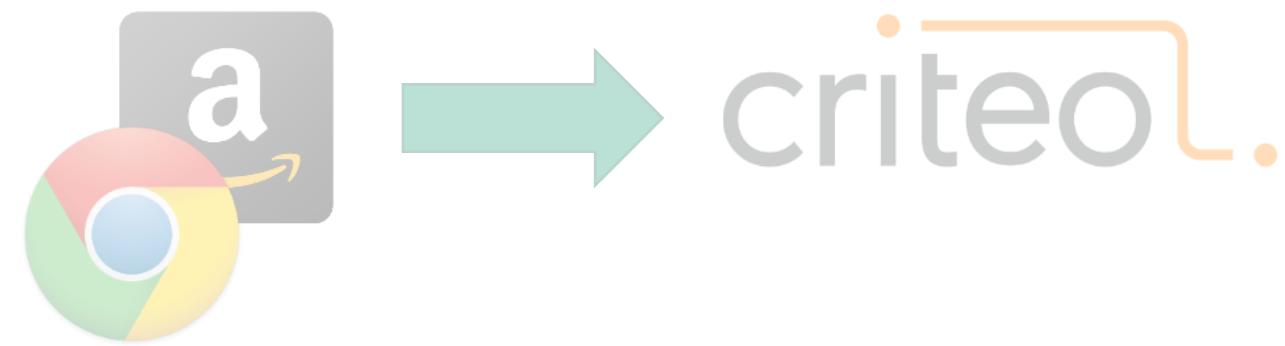
e → d shows that e synched identifiers with d before RTB auction

- d Demand Side Platform (DSP)
- e Exchange (Running Auction)

Example

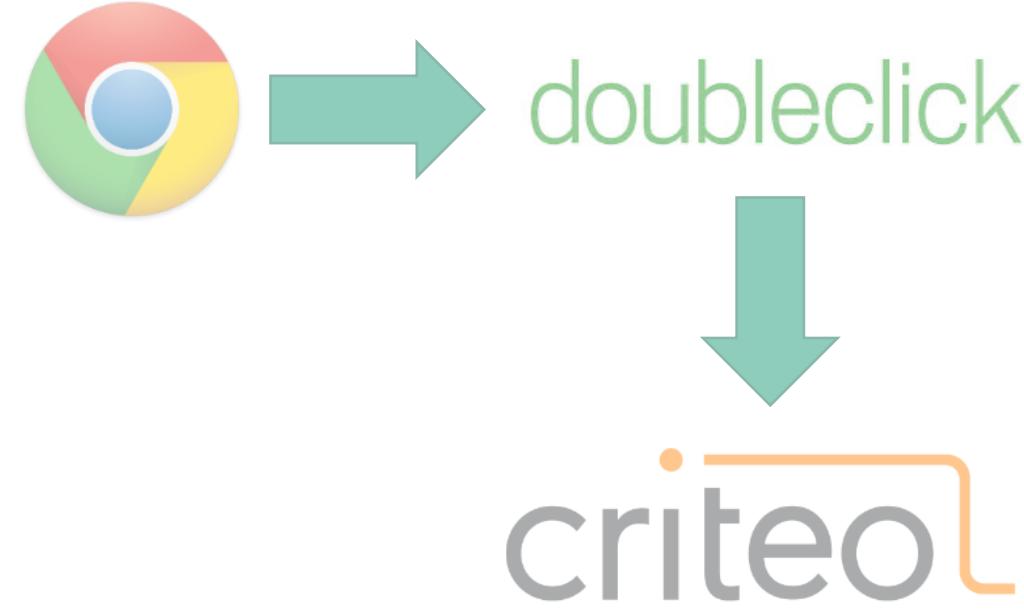
Rule

Shopper-side chain



DSP d must observe us on the shopper side

Anywhere



Transition e → d is where cookie match happens

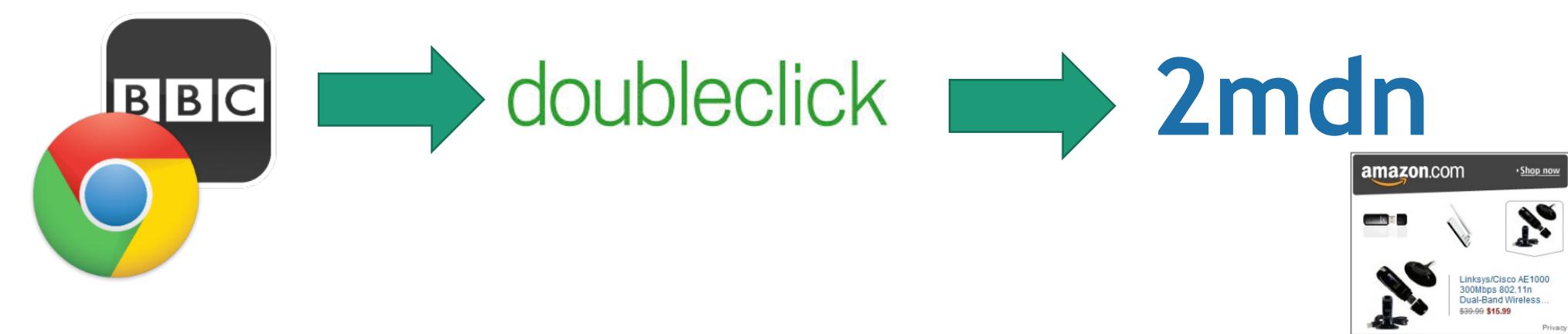
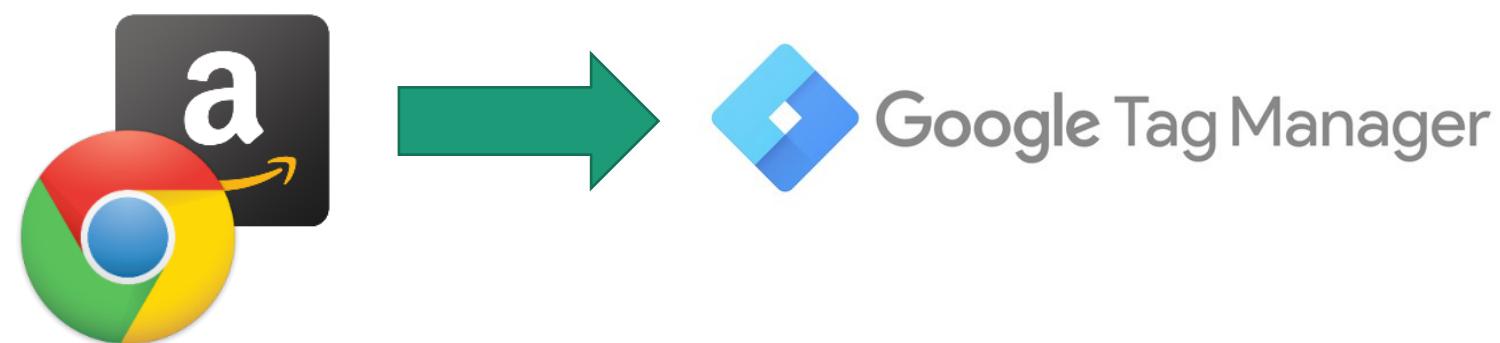
Publisher-side chain



e precedes d, which implies an RTB auction

Latent (Server-Side) Matching

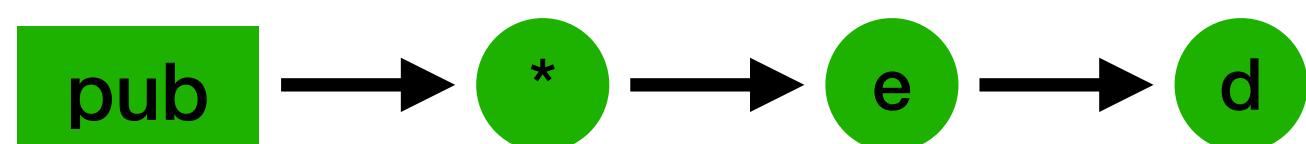
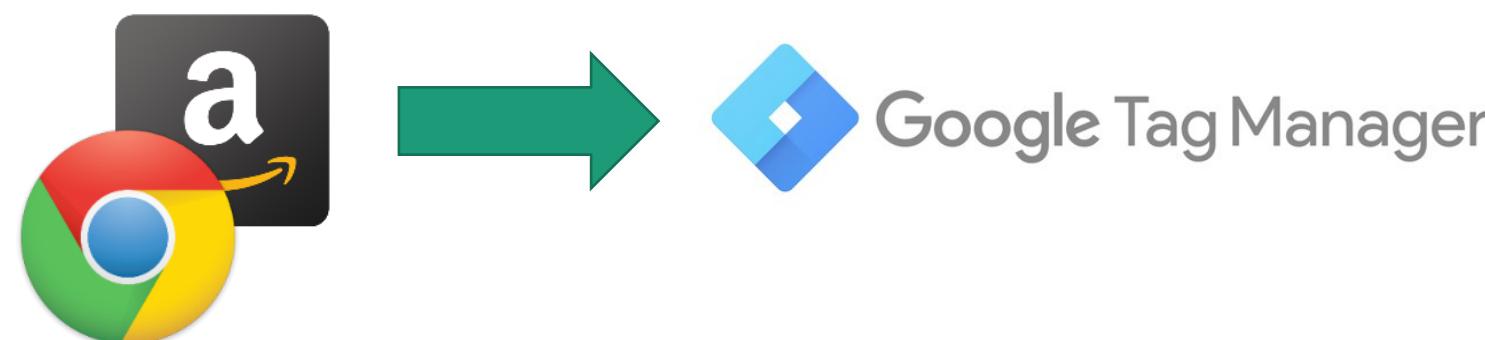
Rule Example



- d Demand Side Platform (DSP)
- e Exchange (Running Auction)

Latent (Server-Side) Matching

Rule Example



- d Demand Side Platform (DSP)
- e Exchange (Running Auction)

Latent (Server-Side) Matching

Rule Example



Google Tag Manager

Shopper-side chain

shop → d'

.....



doubleclick



2mdn

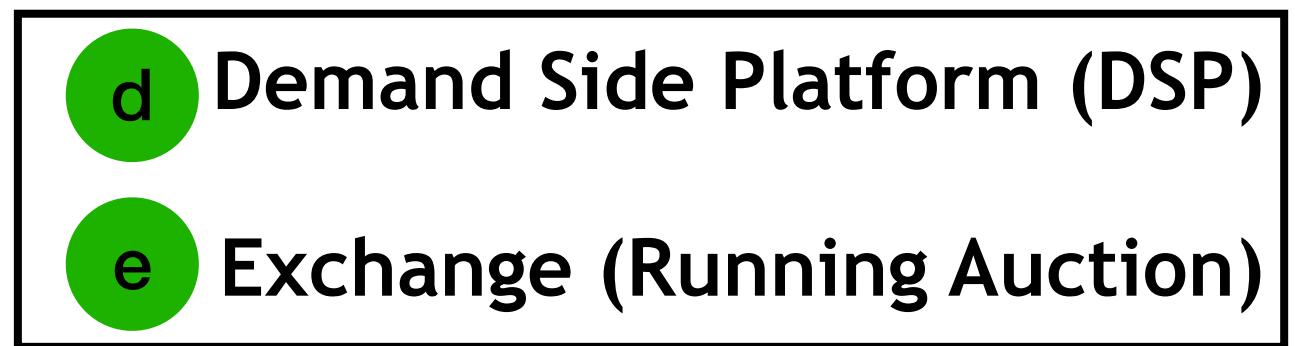
Publisher-side chain

pub → * → e → d



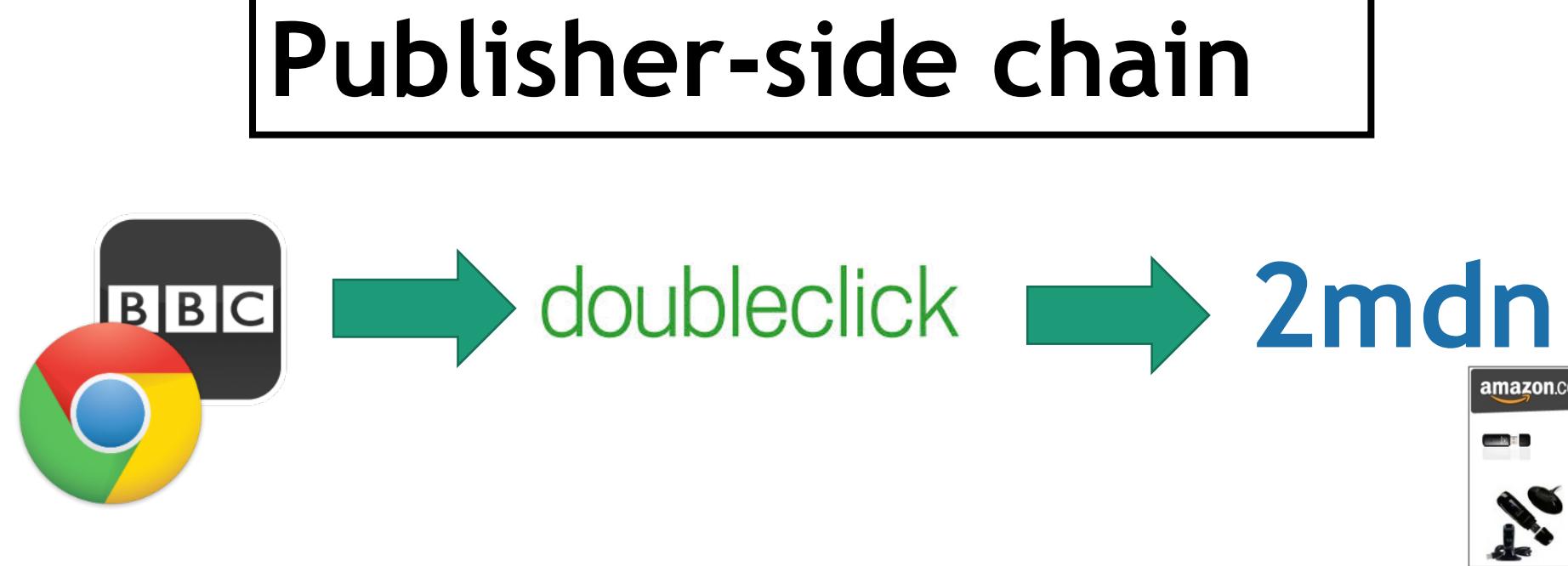
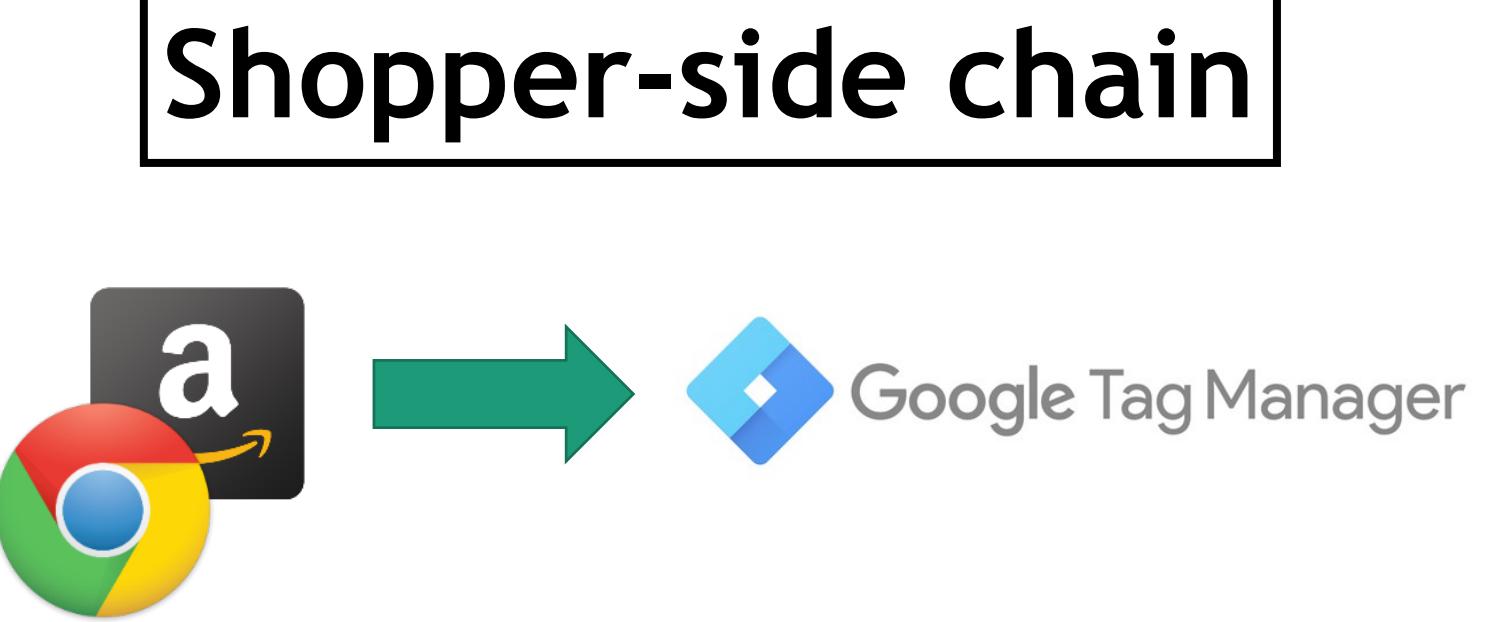
- d Demand Side Platform (DSP)
- e Exchange (Running Auction)

Latent (Server-Side) Matching



Example

Rule



shop → d'

Neither d nor e observes
the user on the shopper
side

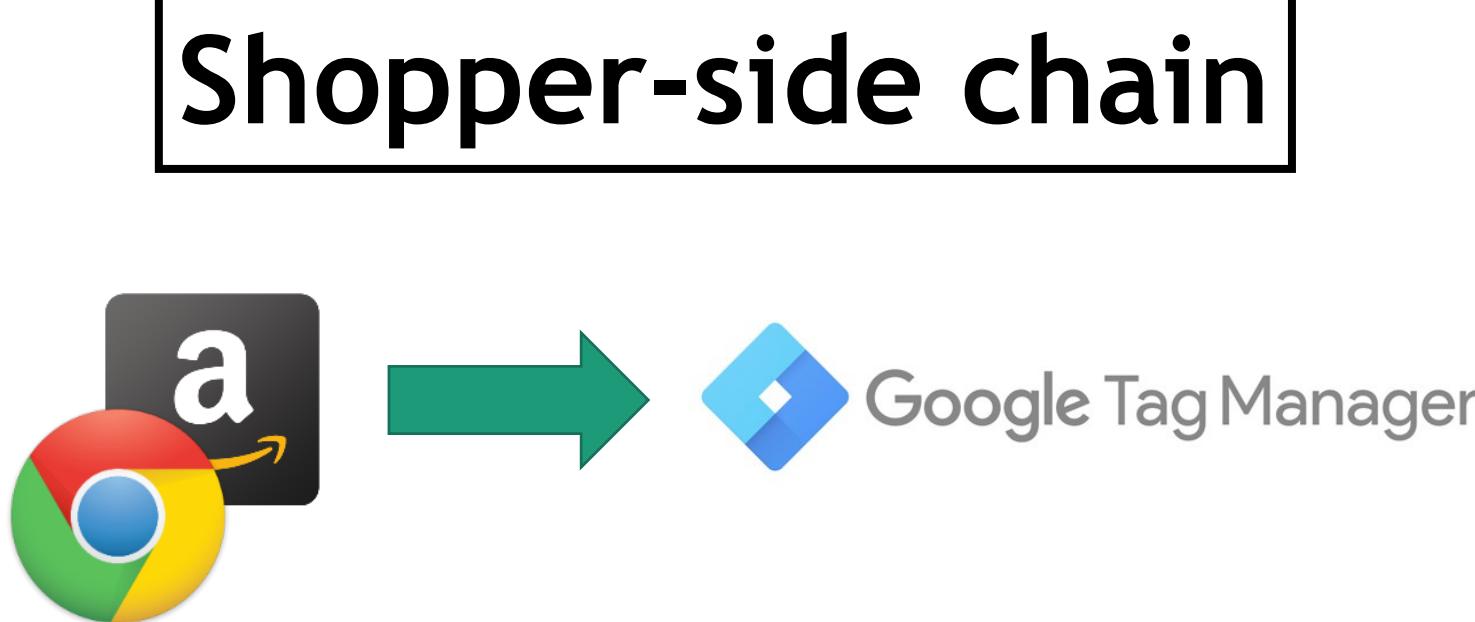
pub → * → e → d

d must receive
information from some
shopper-side tracker

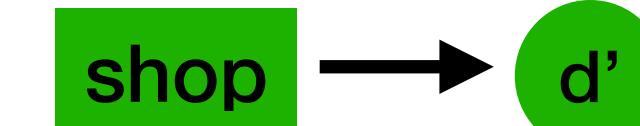
Latent (Server-Side) Matching



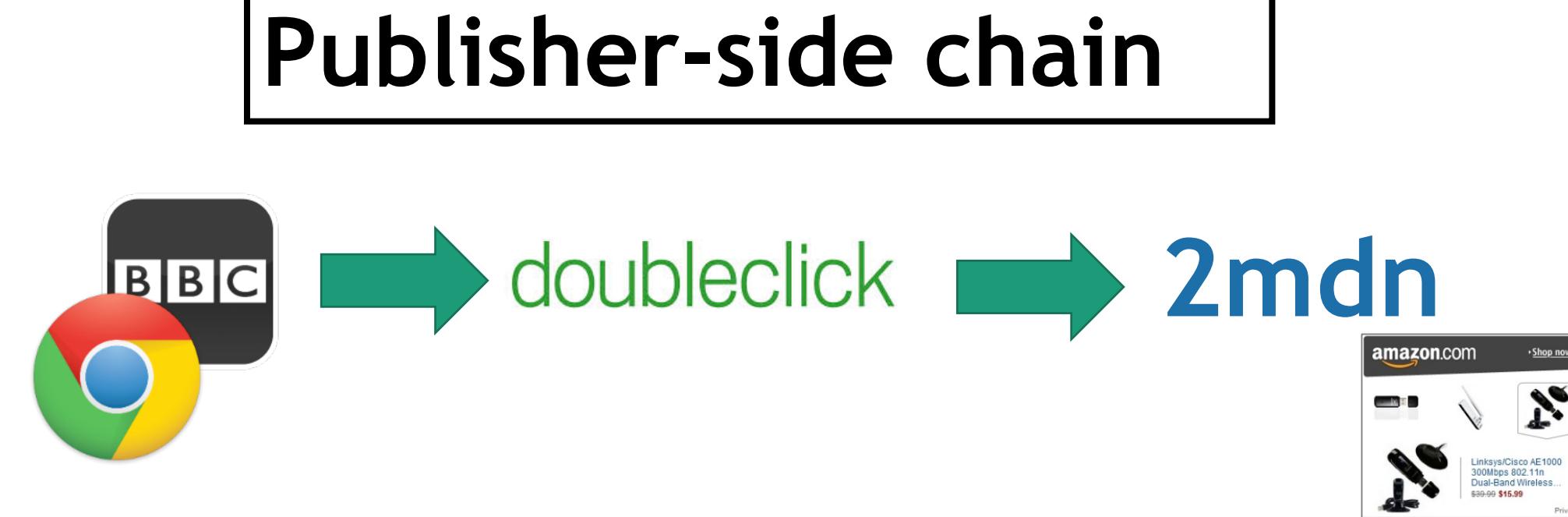
Example



Rule



Neither d nor e observes the user on the shopper side



d must receive information from some shopper-side tracker

d and d' must have shared user identifiers on the server-side

Overall Categorization of Chains

Type	Chains	%
Direct (Trivial) Match	1770	5
Cookie Match	25049	71
Latent Match	5362	15
No Match	775	2

Overall Categorization of Chains

Type	Chains	%
Direct (Trivial) Match	1770	5
Cookie Match	25049	71
Latent Match	5362	15
No Match	775	2

Take away:

Overall Categorization of Chains

Type	Chains	%
Direct (Trivial) Match	1770	5
Cookie Match	25049	71
Latent Match	5362	15
No Match	775	2

Take away:

1. As expected, most retargets are due to cookie matching
 - We detected 200 cookie matching partner pairs

Overall Categorization of Chains

Type	Chains	%
Direct (Trivial) Match	1770	5
Cookie Match	25049	71
Latent Match	5362	15
No Match	775	2

Take away:

1. As expected, most retargets are due to cookie matching
 - We detected 200 cookie matching partner pairs
2. Very small number of chains that cannot be categorized
 - Suggests low false positive rate of AMT image labeling task

Overall Categorization of Chains

Type	Chains	%
Direct (Trivial) Match	1770	5
Cookie Match	25049	71
Latent Match	5362	15
No Match	775	2

Take away:

1. As expected, most retargets are due to cookie matching
 - We detected 200 cookie matching partner pairs
2. Very small number of chains that cannot be categorized
 - Suggests low false positive rate of AMT image labeling task
3. Surprisingly large amount of latent matches!

Clustering Domains

Cluster together domains by “owner”

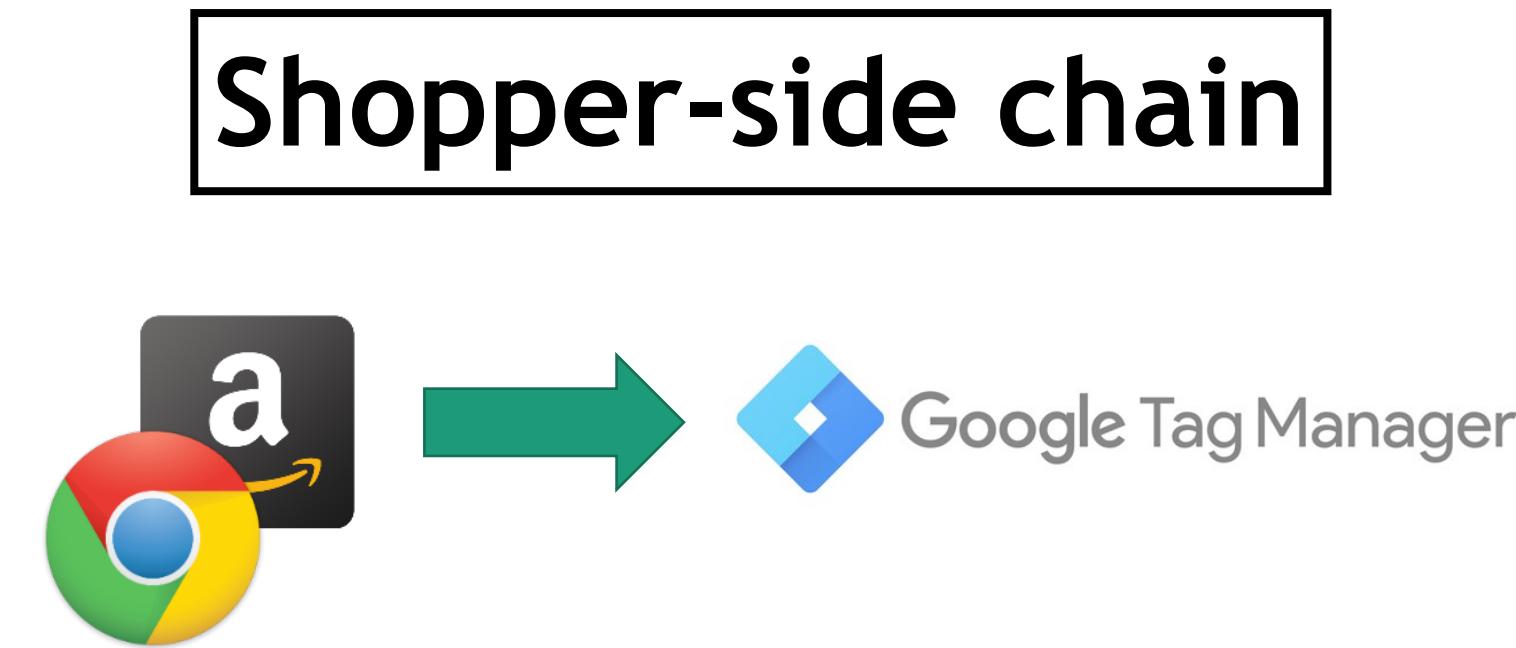
- E.g. google.com, doubleclick.com, googlesyndication.com

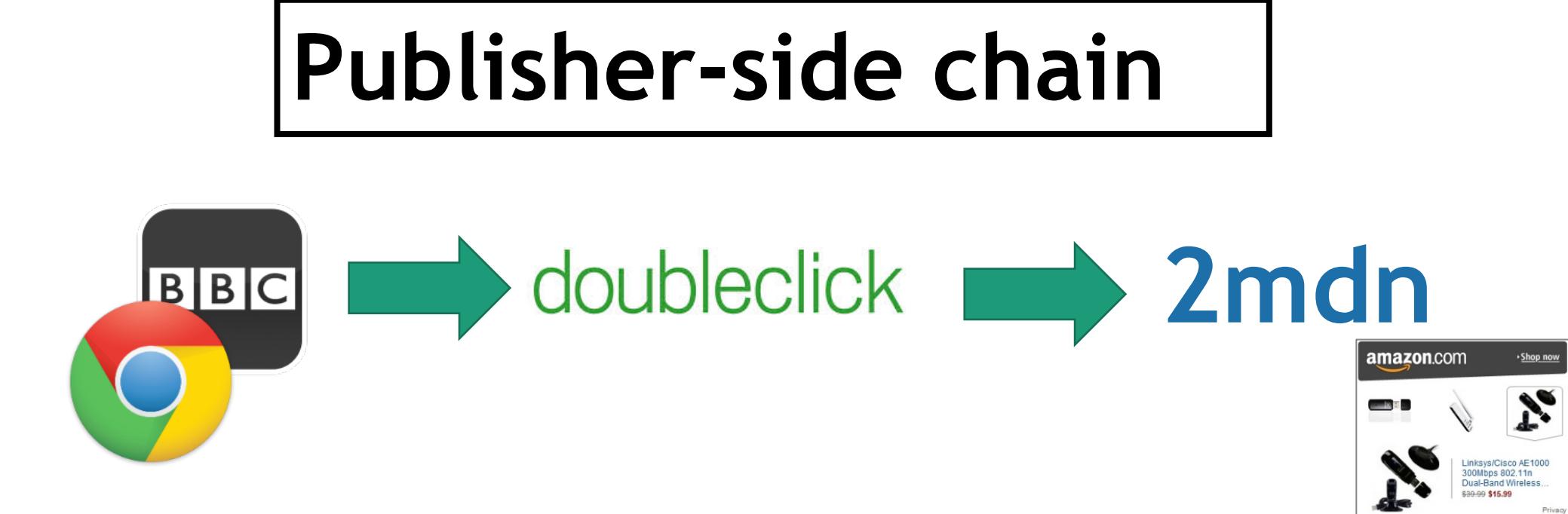
Clustering Domains

Cluster together domains by “owner”

- E.g. google.com, doubleclick.com, googlesyndication.com

Clustered Example

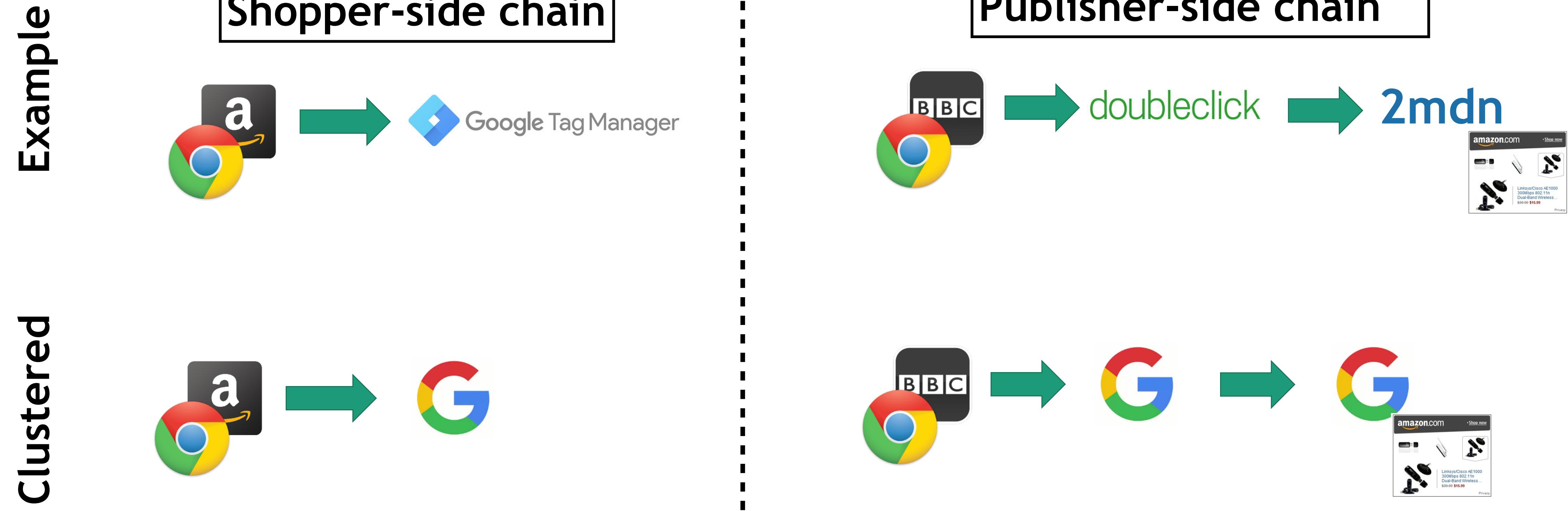




Clustering Domains

Cluster together domains by “owner”

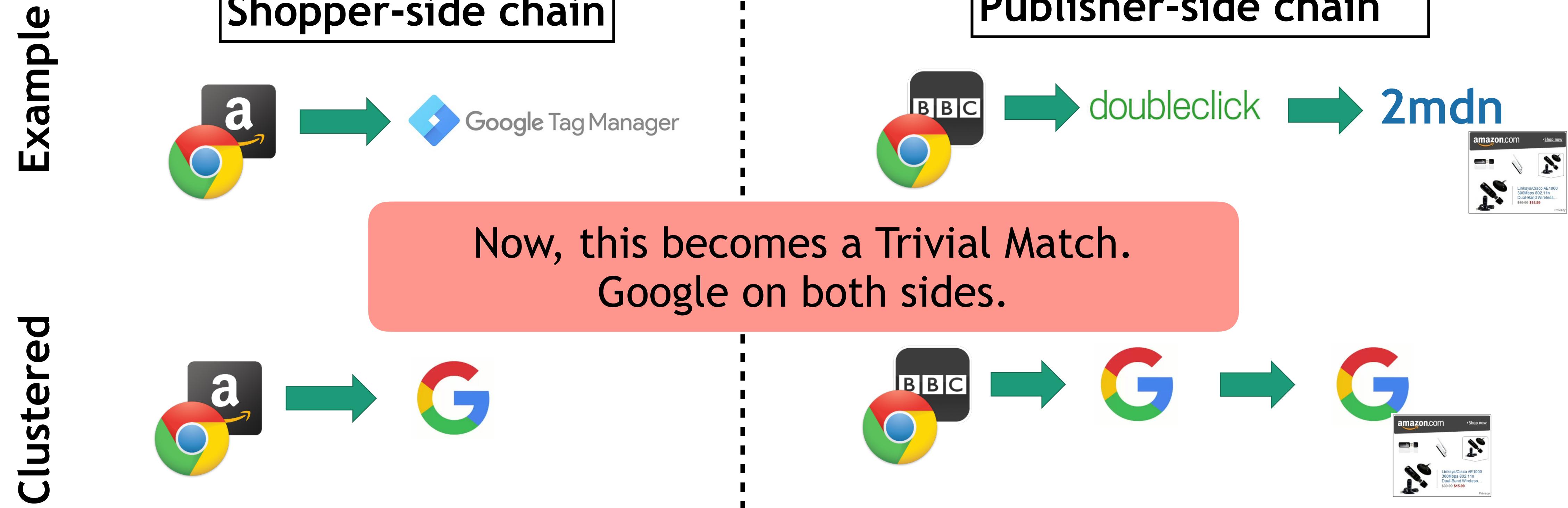
- E.g. google.com, doubleclick.com, googlesyndication.com



Clustering Domains

Cluster together domains by “owner”

- E.g. google.com, doubleclick.com, googlesyndication.com



Overall Categorization of Chains

Type	Raw Chains		Clustered Chains	
	Chains	%	Chains	%
Direct (Trivial) Match	1770	5	8449	24
Cookie Match	25049	71	25873	73
Latent Match	5362	15	343	1
No Match	775	2	183	1

Overall Categorization of Chains

Type	Raw Chains		Clustered Chains	
	Chains	%	Chains	%
Direct (Trivial) Match	1770	5	8449	24
Cookie Match	25049	71	25873	73
Latent Match	5362	15	343	1
No Match	775	2	183	1

Latent matches essentially disappear

- The vast majority of these chains involved Google
- Suggests that Google shares tracking data across their services

Who are the Cookie Matching Partners?

Participant 1		Participant 2	#Chains	#Ads
criteo	↔	googlesyndication	9090	1887
criteo	↔	doubleclick	3610	1144
criteo	↔	adnxss	3263	1066
criteo	↔	rubiconproject	1586	749
criteo	↔	servedbyopenx	707	460
doubleclick	↔	steelhousemedia	362	27
mathtag	↔	mediaforge	360	124
netmng	↔	scene7	267	119
googlesyndication	↔	adsrvr	107	29
rubiconproject	↔	steelhousemedia	86	30
googlesyndication	↔	steelhousemedia	47	22
adtechus	→	adacado	36	18
atwola	→	adacado	32	6
adroll	↔	adnxss	31	8

Who are the Cookie Matching Partners?

Participant 1		Participant 2	#Chains	#Ads
criteo	↔	googlesyndication	9090	1887
criteo	↔	doubleclick	3610	1144
criteo	↔	adnxss	3263	1066
criteo	↔	rubiconproject	1586	749
criteo	↔	servedbyopenx	707	460
doubleclick	↔	steelhousemedia	362	27
mathtag	↔	mediaforge	360	124
netmng	↔	scene7	267	119
googlesyndication	↔	adsrvr	107	29
rubiconproject	↔	steelhousemedia	86	30
googlesyndication	↔	steelhousemedia	47	22
adtechus	→	adacado	36	18
atwola	→	adacado	32	6
adroll	↔	adnxss	31	8

Direction in which
matching happens

Who are the Cookie Matching Partners?

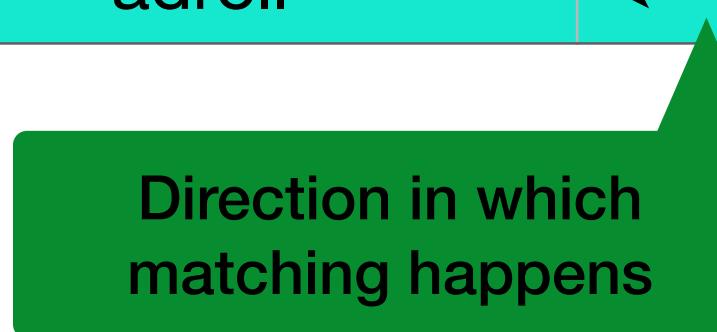
Participant 1		Participant 2	#Chains	#Ads
criteo	↔	googlesyndication	9090	1887
criteo	↔	doubleclick	3610	1144
criteo	↔	adnxss	3263	1066
criteo	↔	rubiconproject	1586	749
criteo	↔	servedbyopenx	707	460
doubleclick	↔	steelhousemedia	362	27
mathtag	↔	mediaforge	360	124
netmng	↔	scene7	267	119
googlesyndication	↔	adsrvr	107	29
rubiconproject	↔	steelhousemedia	86	30
googlesyndication	↔	steelhousemedia	47	22
adtechus	→	adacado	36	18
atwola	→	adacado	32	6
adroll	↔	adnxss	31	8

Direction in which
matching happens

Who are the Cookie Matching Partners?

Participant 1	Participant 2	#Chains	#Ads	Heuristics (Prior Work)
criteo	← →	googlesyndication	9090	1887 ← → P
criteo	← →	doubleclick	3610	1144 → E, P ← DC, P
criteo	← →	adnxns	3263	1066 ← → E, P
criteo	← →	rubiconproject	1586	749 ← → E, P
criteo	← →	servedbyopenx	707	460 ← → P
doubleclick	← →	steelhousemedia	362	27 → P ← E, P
mathtag	← →	mediaforge	360	124 ← → E, P
netmng	← →	scene7	267	119 → E ← ?
googlesyndication	← →	adsrvr	107	29 ← → P
rubiconproject	← →	steelhousemedia	86	30 ← → E
googlesyndication	← →	steelhousemedia	47	22 ?
adtechus	→	adacado	36	18 ?
atwola	→	adacado	32	6 ?
adroll	← →	adnxns	31	8 ?

Direction in which matching happens



Heuristics Key (used by prior work)

E - exact identifier value shared

P - special URL parameters
e.g. cookie_sync, user_sync

DC - DoubleClick URL parameters
e.g. google_gid, google_nid

? - Match missed by heuristics

Who are the Cookie Matching Partners?

Participant 1	Participant 2	#Chains	#Ads	Heuristics (Prior Work)
criteo	↔ ↔	googlesyndication	9090	1887 ← → P
criteo	↔ ↔	doubleclick	3610	1144 → E, P ← DC, P
criteo	↔ ↔	adnxns	3263	1066 ← → E, P
criteo	↔ ↔	rubiconproject	1586	749 ← → E, P
criteo	↔ ↔	servedbyopenx	707	460 ← → P
doubleclick	↔ ↔	steelhousemedia	362	27 → P ← E, P
mathtag	↔ ↔	mediaforge	360	124 ← → E, P
netmng	↔ ↔	scene7	267	119 → E ← ?
googlesyndication	↔ ↔	adsrvr	107	29 ← → P
rubiconproject	↔ ↔	steelhousemedia	86	30 ← → E
googlesyndication	↔ ↔	steelhousemedia	47	22 ?
adtechus	→	adacado	36	18 ?
atwola	→	adacado	32	6 ?
adroll	↔ ↔	adnxns	31	8 ?

Heuristics Key (used by prior work)

E - exact identifier value shared

P - special URL parameters
e.g. cookie_sync, user_sync

DC - DoubleClick URL parameters
e.g. google_gid, google_nid

? - Match missed by heuristics

Direction in which
matching happens

Who are the Cookie Matching Partners?

Participant 1	Participant 2	#Chains	#Ads	Heuristics (Prior Work)
criteo	↔ ↔	googlesyndication	9090	1887
criteo	↔ ↔	doubleclick	3610	1144 → E, P ← DC, P
criteo	↔ ↔	adnxs	3263	1066 ← → E, P
criteo	↔ ↔	rubiconproject	1586	749 ← → E, P
criteo	↔ ↔	servedbyopenx	707	460 ← → P
doubleclick	↔ ↔	steelhousemedia	362	27 → P ← E, P
mathtag	↔ ↔	mediaforge	360	124 ← → E, P
netmng	↔ ↔	scene7	267	119 → E ← ?
googlesyndication	↔ ↔	adsrvr	107	29 ← → P
rubiconproject	↔ ↔	steelhousemedia	86	30 ← → E
googlesyndication	↔ ↔	steelhousemedia	47	22 ?
adtechus	→	adacado	36	18 ?
atwola	→	adacado	32	6 ?
adroll	↔ ↔	adnxs	31	8 ?

Direction in which matching happens

Heuristics Key (used by prior work)

E - exact identifier value shared

P - special URL parameters
e.g. cookie_sync, user_sync

DC - DoubleClick URL parameters
e.g. google_gid, google_nid

? - Match missed by heuristics

31% of cookie matching partners were missed by prior heuristics

Summary so far...

Novel methodology to detect information flows between ad exchanges

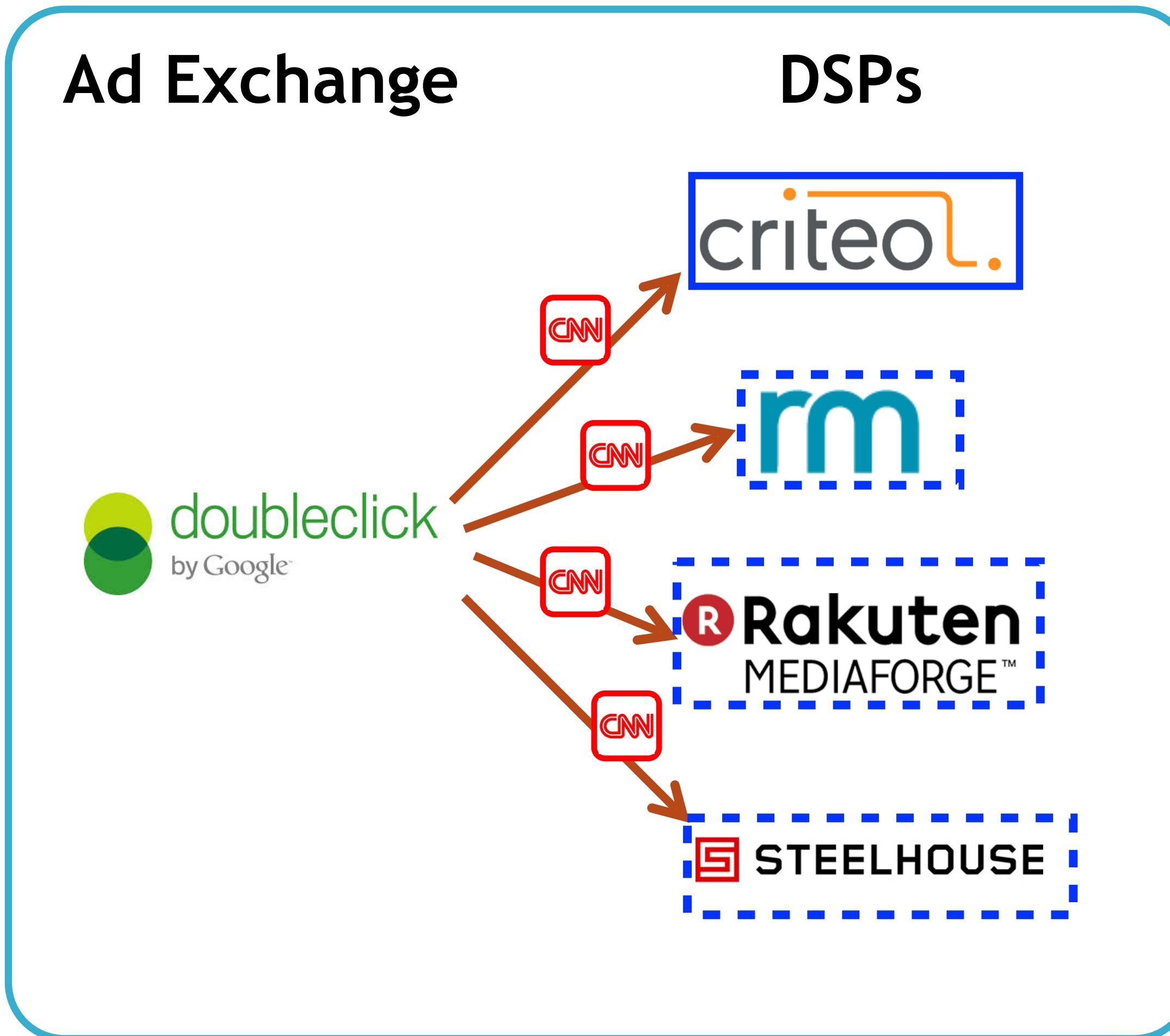
- Controlled methodology enables us to make causal inferences
- Defeats obfuscation attempts, since we do not rely on HTTP headers
- Detects client- and server-side information flows

Identified 200 cookie matching partner pairs

- 31% would have been missed by prior techniques

Close Collaboration Among Ad Networks

RTB Auction



Takeaway #1

- Cookie Matching is a pre-requisite for RTB
- This leads to close collaboration among ad companies

Takeaway #2

- In RTB, all bid-accepting DSP partners view user impression
- The losing parties in RTB also learn this information

Outline

Background & Motivation

Limitations of Prior Work

Detecting Information Sharing

- Using Retargeted Ads

Propagation of Tracking Data

- Modeling via Graph

Ongoing Work

- Cross-device Tracking
- Automation of Adblocking Filter Lists

Research Plan

Diffusion of User Tracking Data in the Advertising Ecosystem

Goal: Model the Diffusion of Impressions in the Advertising Ecosystem

Taking RTB into account

Diffusion of User Tracking Data in the Advertising Ecosystem

Goal: Model the Diffusion of Impressions in the Advertising Ecosystem

Taking RTB into account

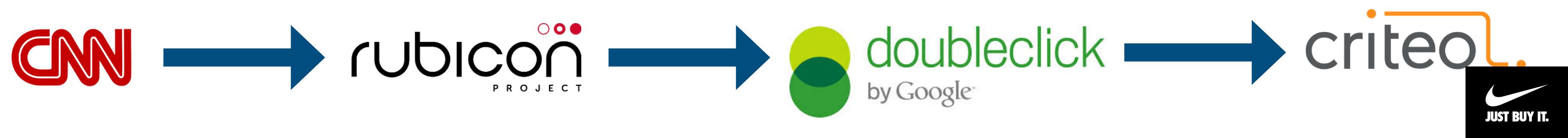
Key Questions:

1. What fraction of user impressions are viewed by ad companies?
2. How much ad and tracker blocking extensions help?

Key Terms:

1. **Impressions:** Page Visits
2. **Publishers:** First party websites visited by users (e.g. cnn, bbc, espn)
3. **A&A:** Advertising and Analytics related companies / domains

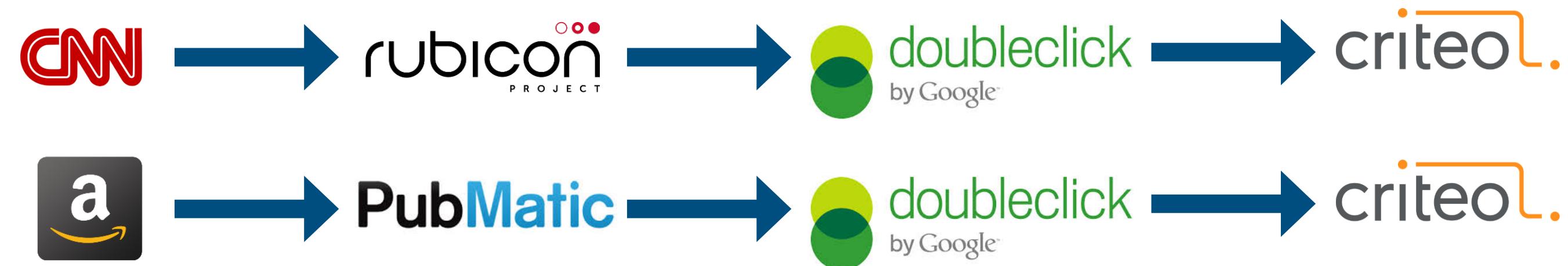
Recap - Inclusion Chains



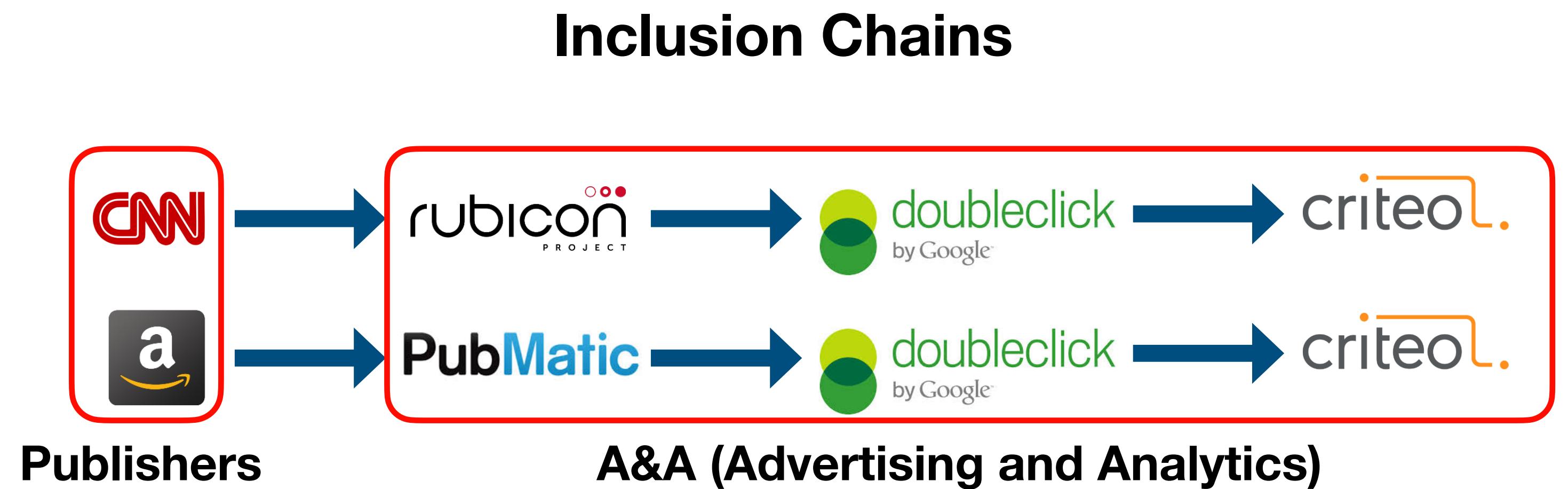
From Chains to Graph

From Chains to Graph

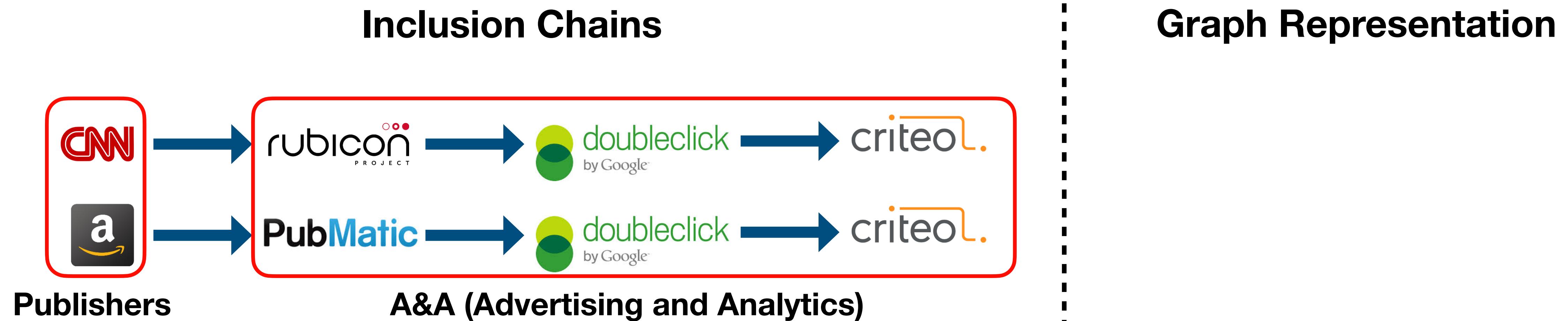
Inclusion Chains



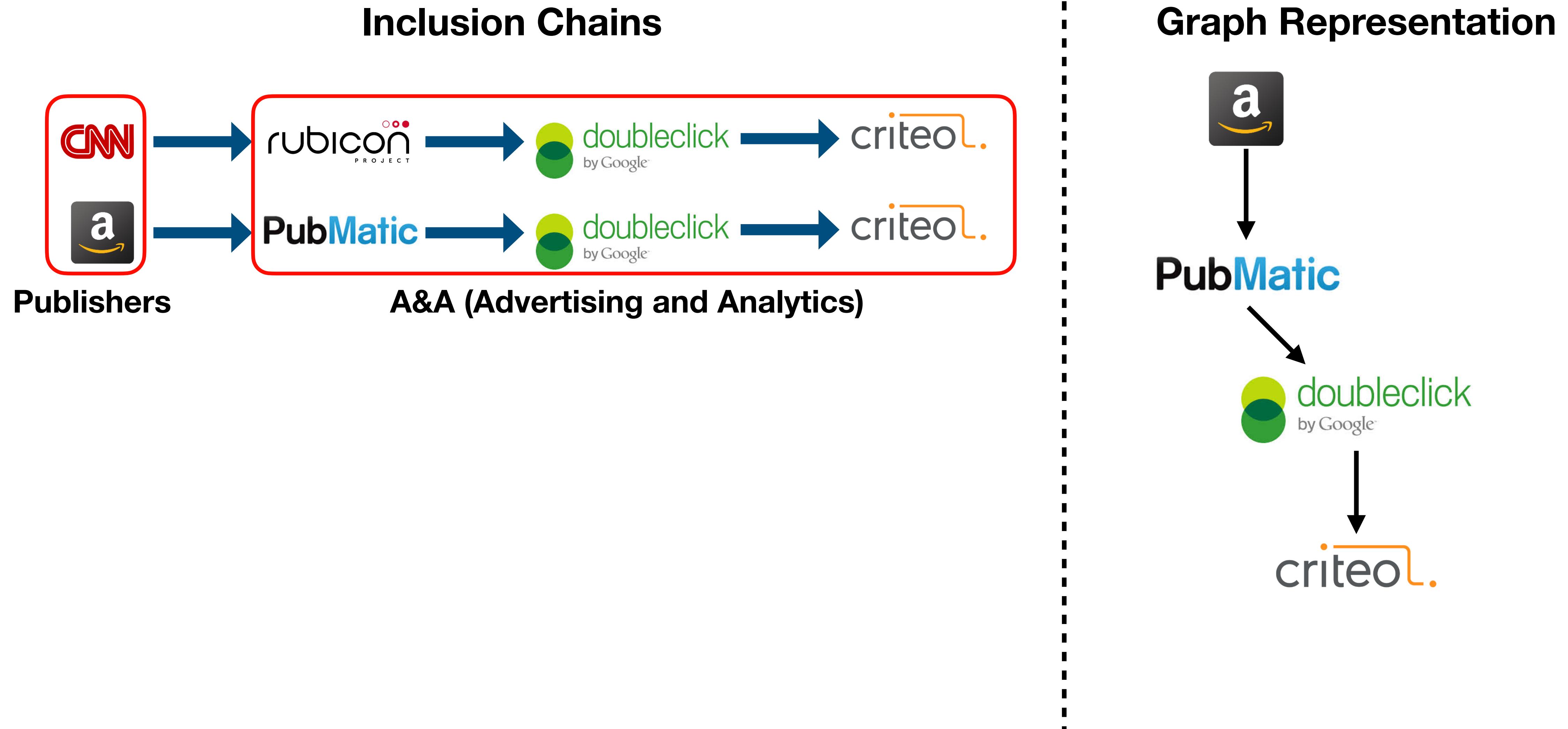
From Chains to Graph



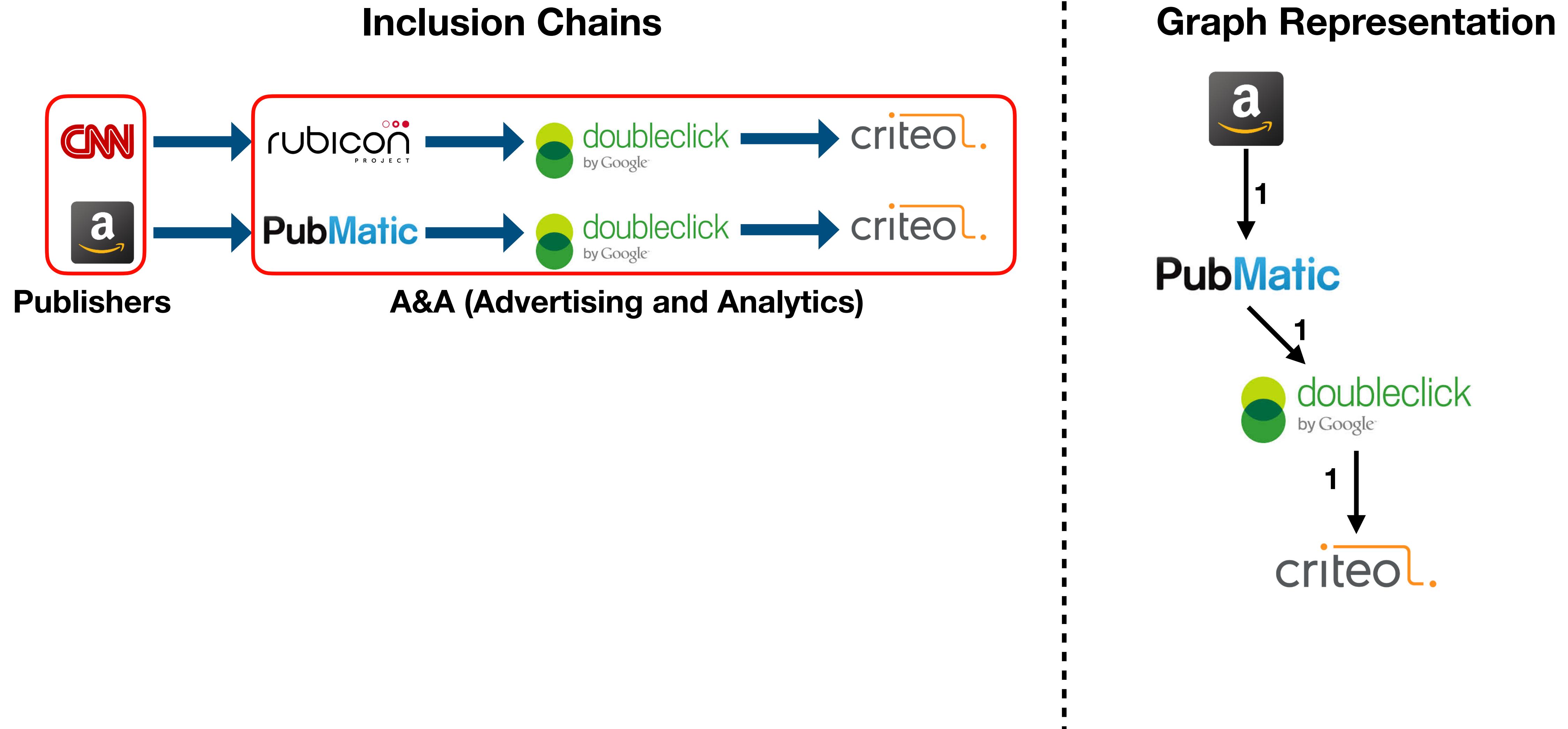
From Chains to Graph



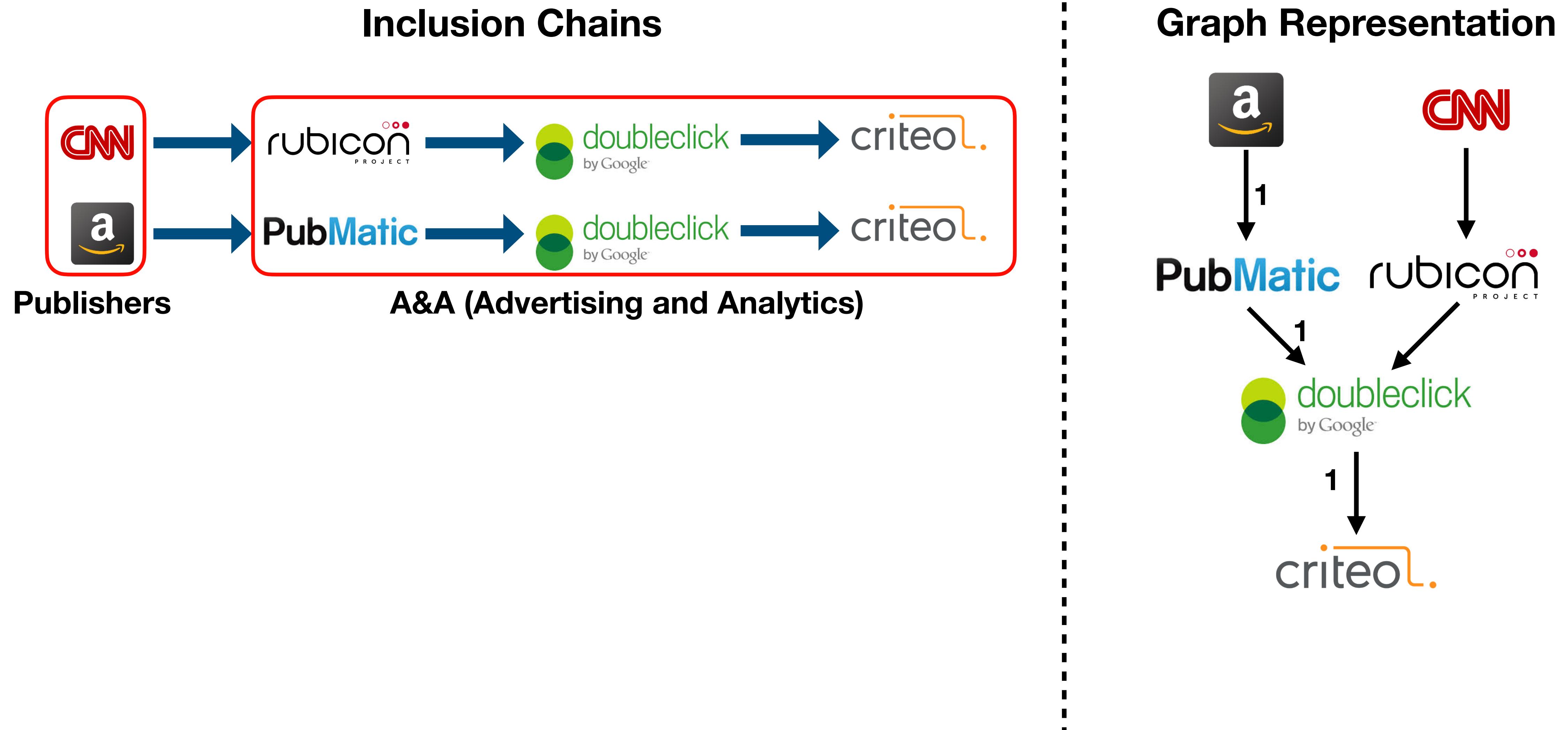
From Chains to Graph



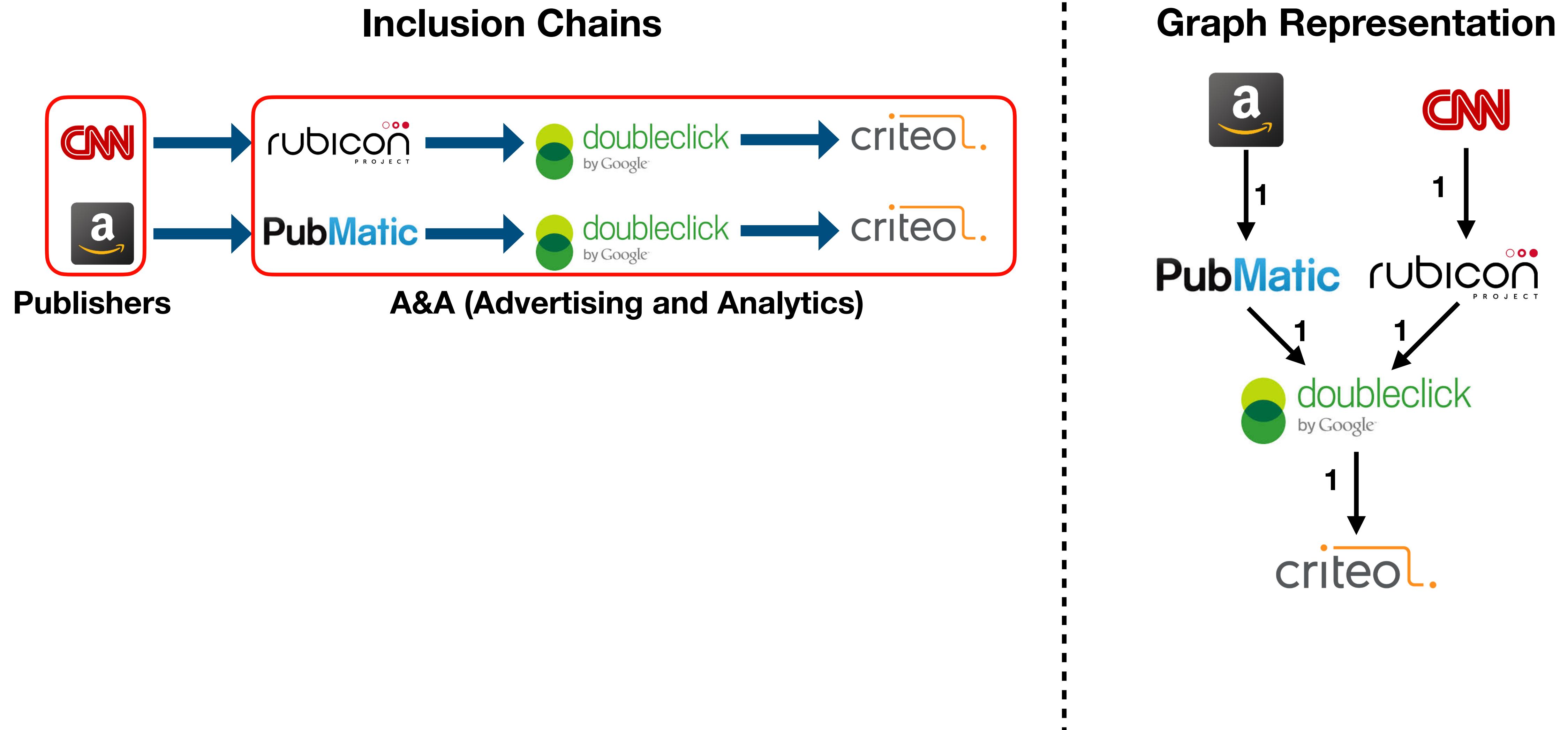
From Chains to Graph



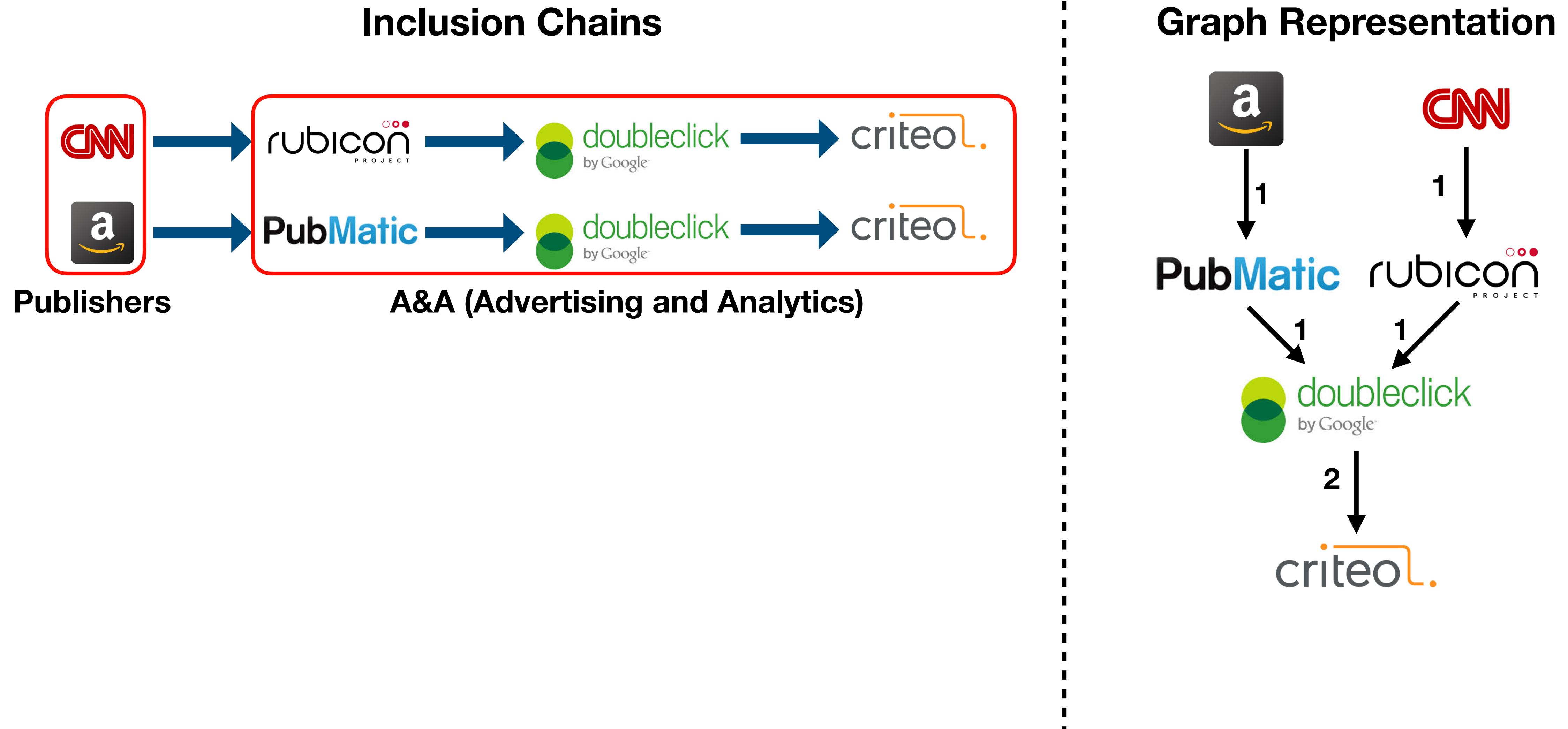
From Chains to Graph



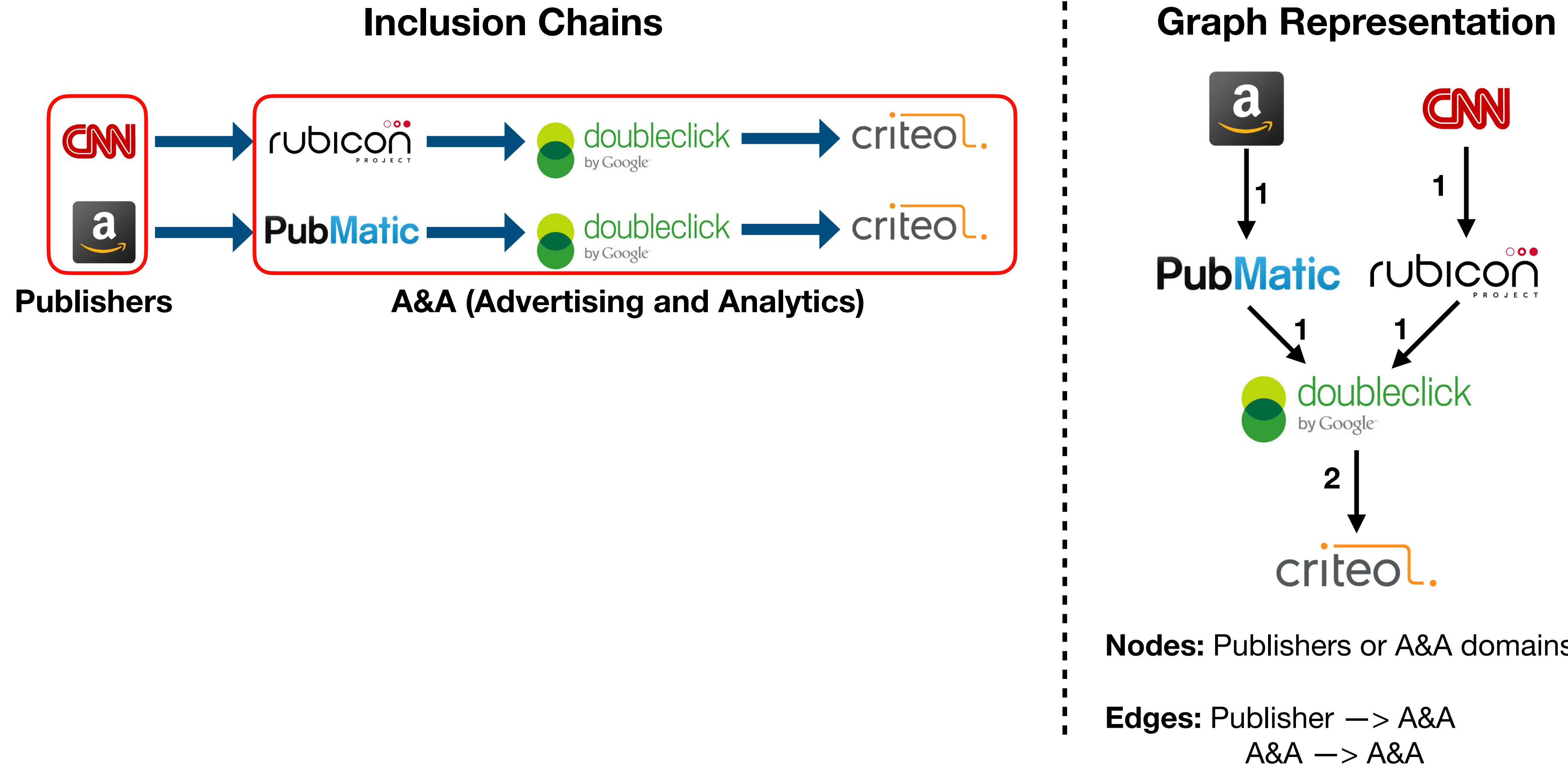
From Chains to Graph



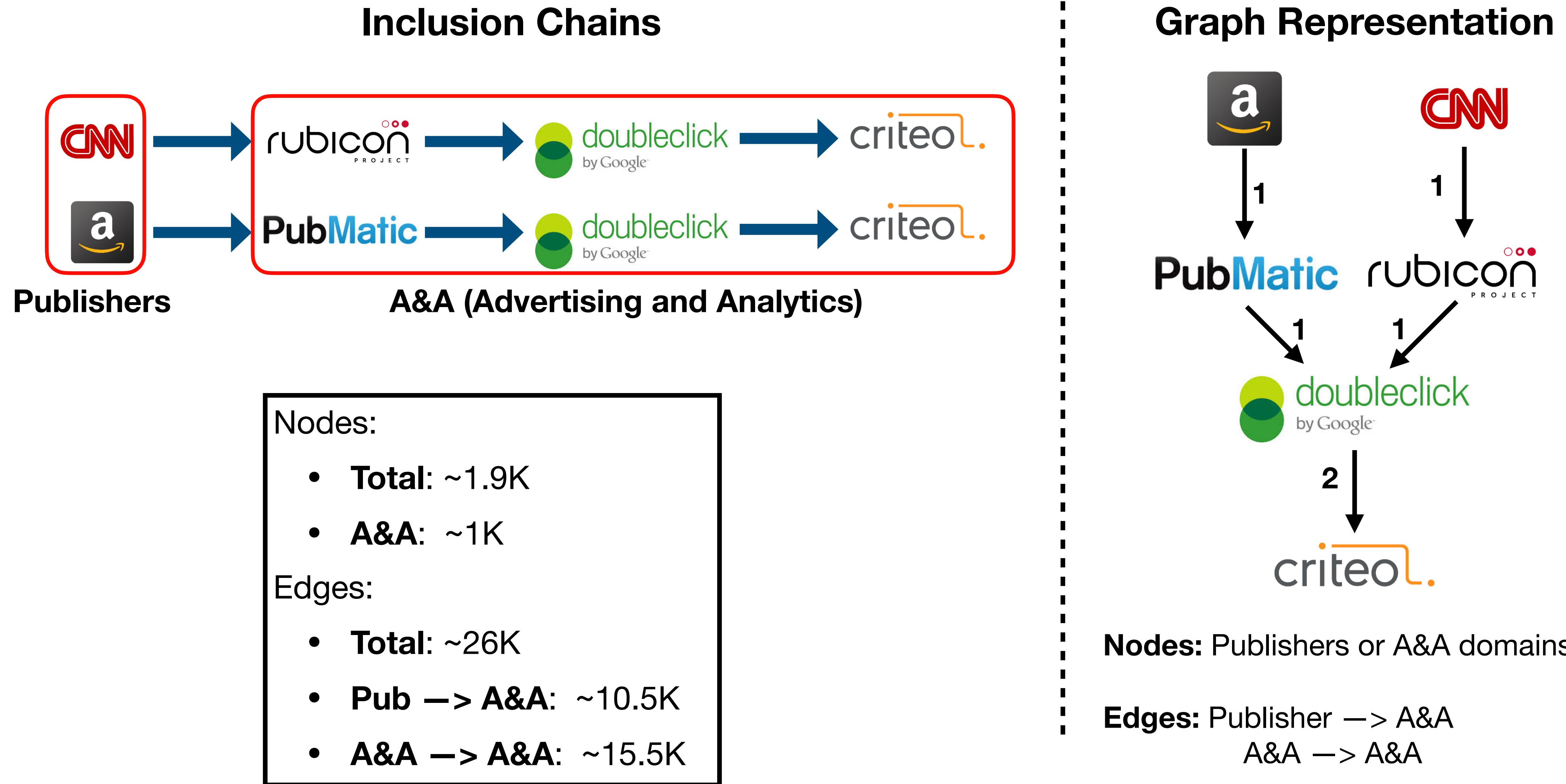
From Chains to Graph



From Chains to Graph



From Chains to Graph



Some High-level Properties of the Graph

Some High-level Properties of the Graph

- The graph is very dense.

Some High-level Properties of the Graph

- The graph is very dense.
- No distinct communities
 - Web is not necessarily balkanized into distinct groups

Some High-level Properties of the Graph

- The graph is very dense.
- No distinct communities
 - Web is not necessarily balkanized into distinct groups
- Expected top nodes with PageRank and Betweenness Centrality



A major goal is to model
diffusion of tracking data

Simulation Setup for Propagation of Tracking Data

[1]. Burken et al. User centric walk: An integrated approach for modeling the browsing behavior of users on the web. ASS 2005

Simulation Setup for Propagation of Tracking Data

We simulate browsing traces for 200 users using method from [1].

[1]. Burken et al. User centric walk: An integrated approach for modeling the browsing behavior of users on the web. ASS 2005

Simulation Setup for Propagation of Tracking Data

We simulate browsing traces for 200 users using method from [1].

1. User generates an impression on N selected publishers.

[1]. Burken et al. User centric walk: An integrated approach for modeling the browsing behavior of users on the web. ASS 2005

Simulation Setup for Propagation of Tracking Data

We simulate browsing traces for 200 users using method from [1].

1. User generates an impression on N selected publishers.
2. Impressions are forwarded to A&A domains via:

A. **Direct Propagation:** 

- Present on publisher or won RTB auction. **Observable** (goes through the browser)

B. **Indirect Propagation:** 

- A&A domains learn impressions through RTB participation. **Non-observable**

[1]. Burken et al. User centric walk: An integrated approach for modeling the browsing behavior of users on the web. ASS 2005

Simulation Setup for Propagation of Tracking Data

We simulate browsing traces for 200 users using method from [1].

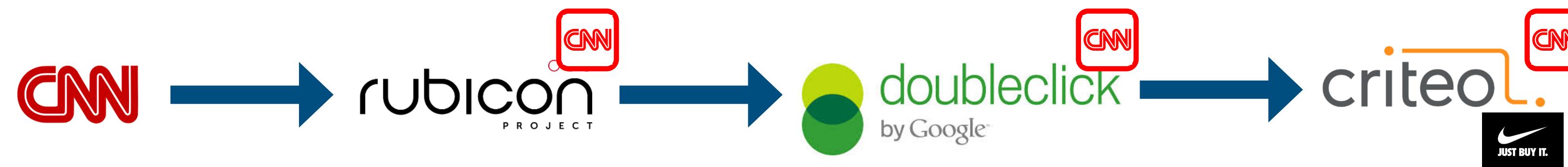
1. User generates an impression on N selected publishers.
2. Impressions are forwarded to A&A domains via:

A. **Direct Propagation:** 

- Present on publisher or won RTB auction. **Observable** (goes through the browser)

B. **Indirect Propagation:** 

- A&A domains learn impressions through RTB participation. **Non-observable**



[1]. Burken et al. User centric walk: An integrated approach for modeling the browsing behavior of users on the web. ASS 2005

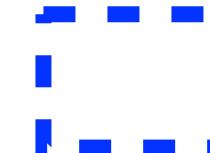
Simulation Setup for Propagation of Tracking Data

We simulate browsing traces for 200 users using method from [1].

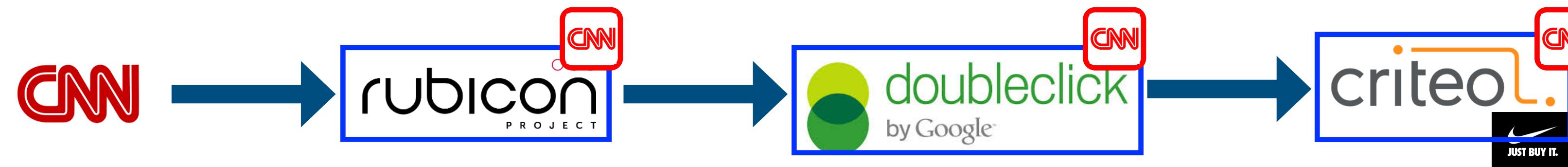
1. User generates an impression on N selected publishers.
2. Impressions are forwarded to A&A domains via:

A. **Direct Propagation:** 

- Present on publisher or won RTB auction. **Observable** (goes through the browser)

B. **Indirect Propagation:** 

- A&A domains learn impressions through RTB participation. **Non-observable**



[1]. Burken et al. User centric walk: An integrated approach for modeling the browsing behavior of users on the web. ASS 2005

Simulation Setup for Propagation of Tracking Data

We simulate browsing traces for 200 users using method from [1].

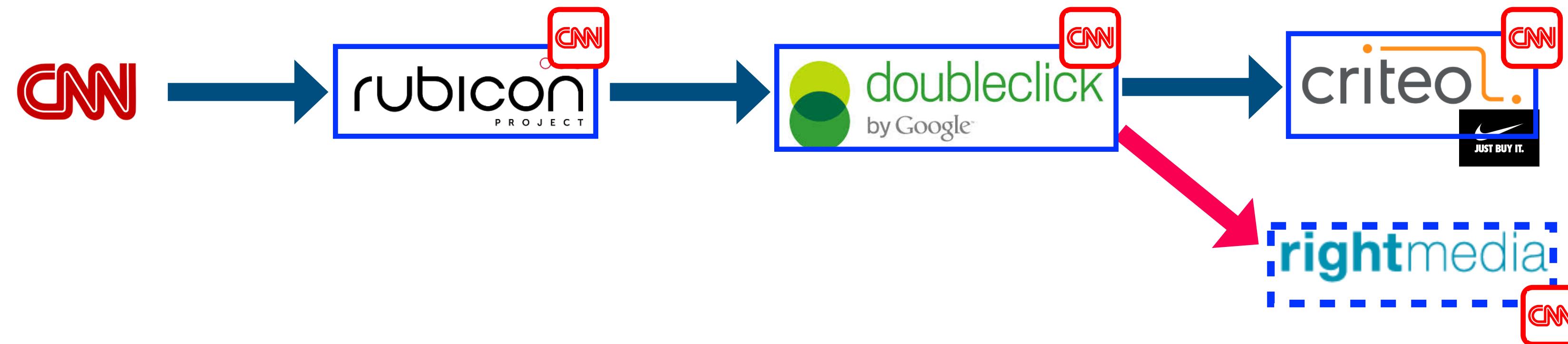
1. User generates an impression on N selected publishers.
2. Impressions are forwarded to A&A domains via:

A. **Direct Propagation:** 

- Present on publisher or won RTB auction. **Observable** (goes through the browser)

B. **Indirect Propagation:** 

- A&A domains learn impressions through RTB participation. **Non-observable**



[1]. Burken et al. User centric walk: An integrated approach for modeling the browsing behavior of users on the web. ASS 2005

Simulation Setup for Propagation of Tracking Data

We simulate browsing traces for 200 users using method from [1].

1. User generates an impression on N selected publishers.

2. Impressions are forwarded to A&A domains via:

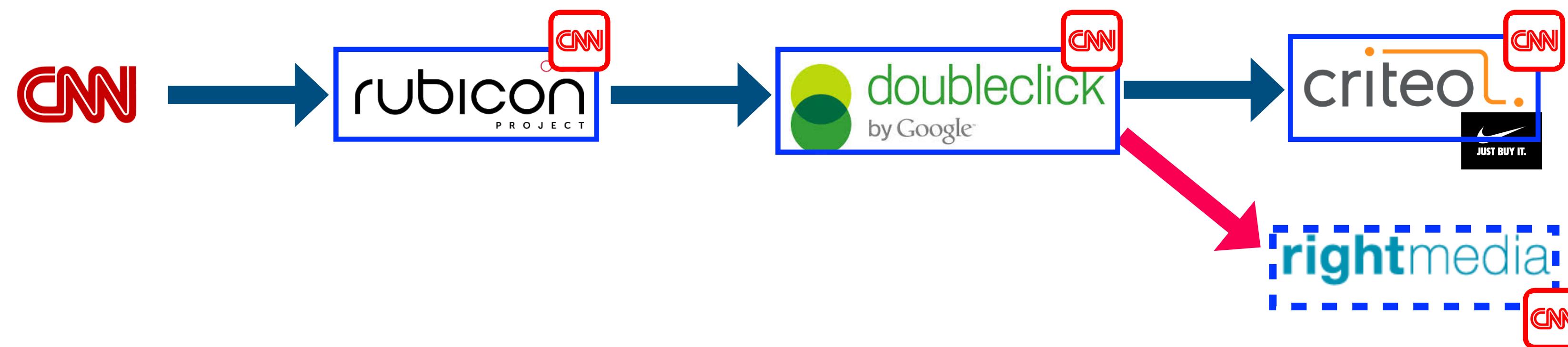
A. **Direct Propagation:** 

- Present on publisher or won RTB auction. **Observable** (goes through the browser)

B. **Indirect Propagation:** 

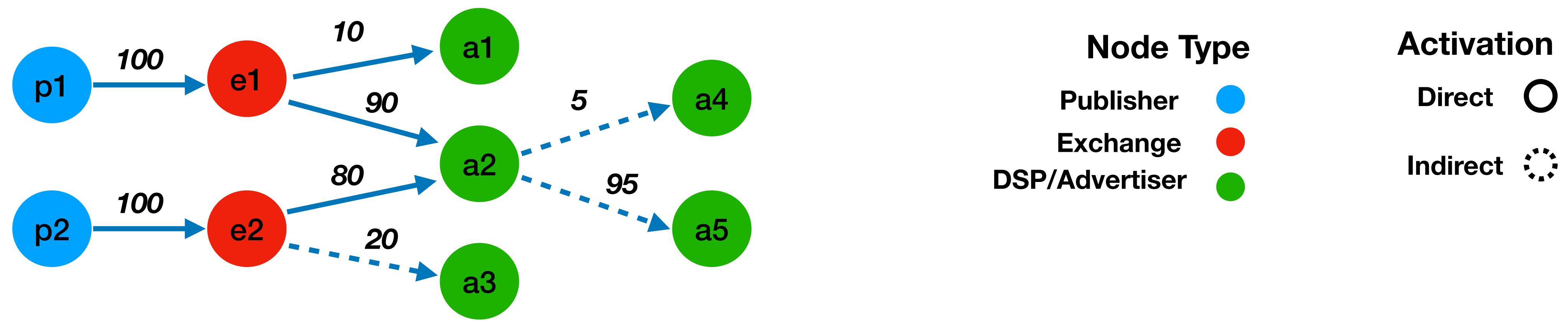
- A&A domains learn impressions through RTB participation. **Non-observable**

3. RTB winner is decided based on probability (function of edge weights).

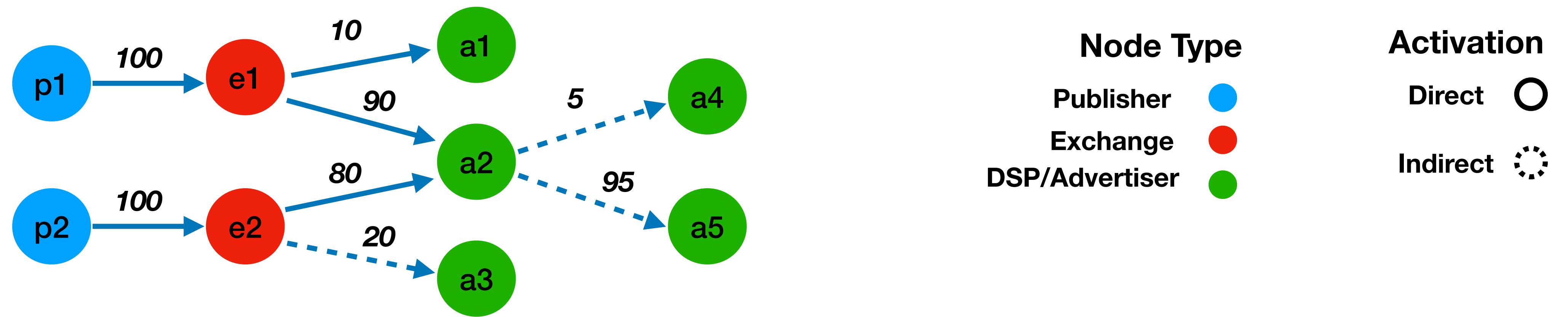


[1]. Burken et al. User centric walk: An integrated approach for modeling the browsing behavior of users on the web. ASS 2005

Simulation Example

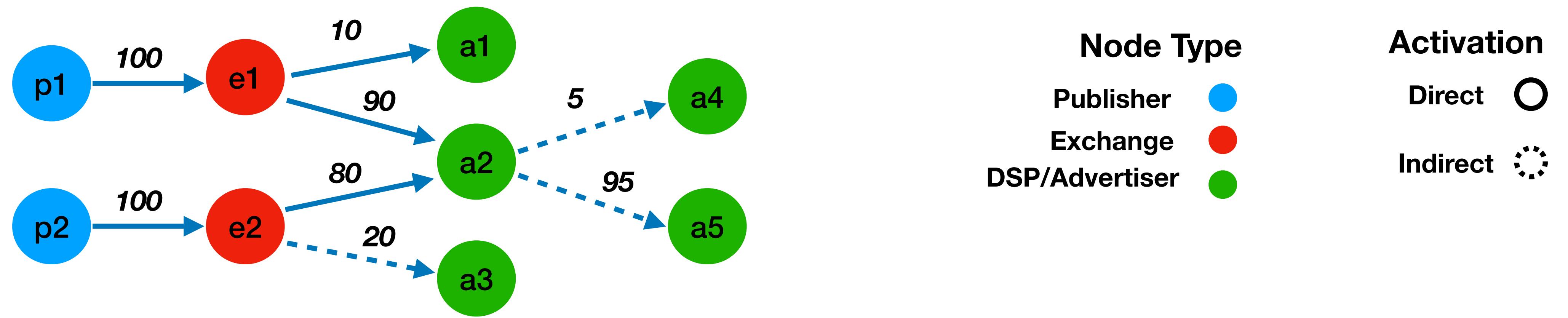


Simulation Example



We manually select **37 exchanges** which are allowed to forward **indirect** impressions to solicit bids during RTB

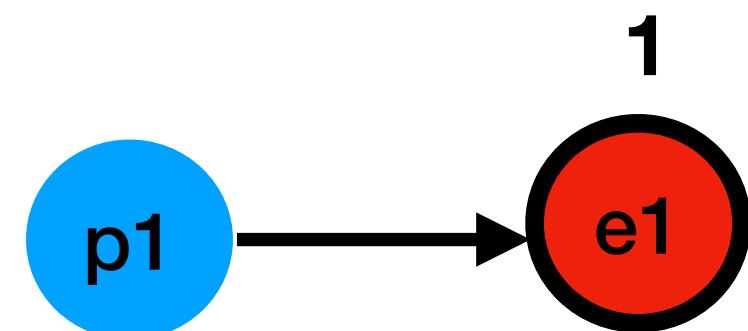
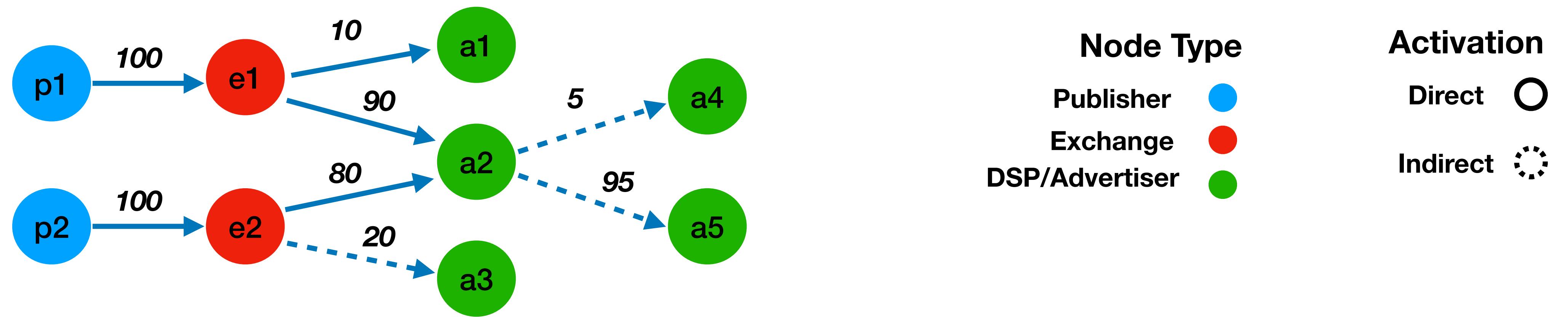
Simulation Example



p1

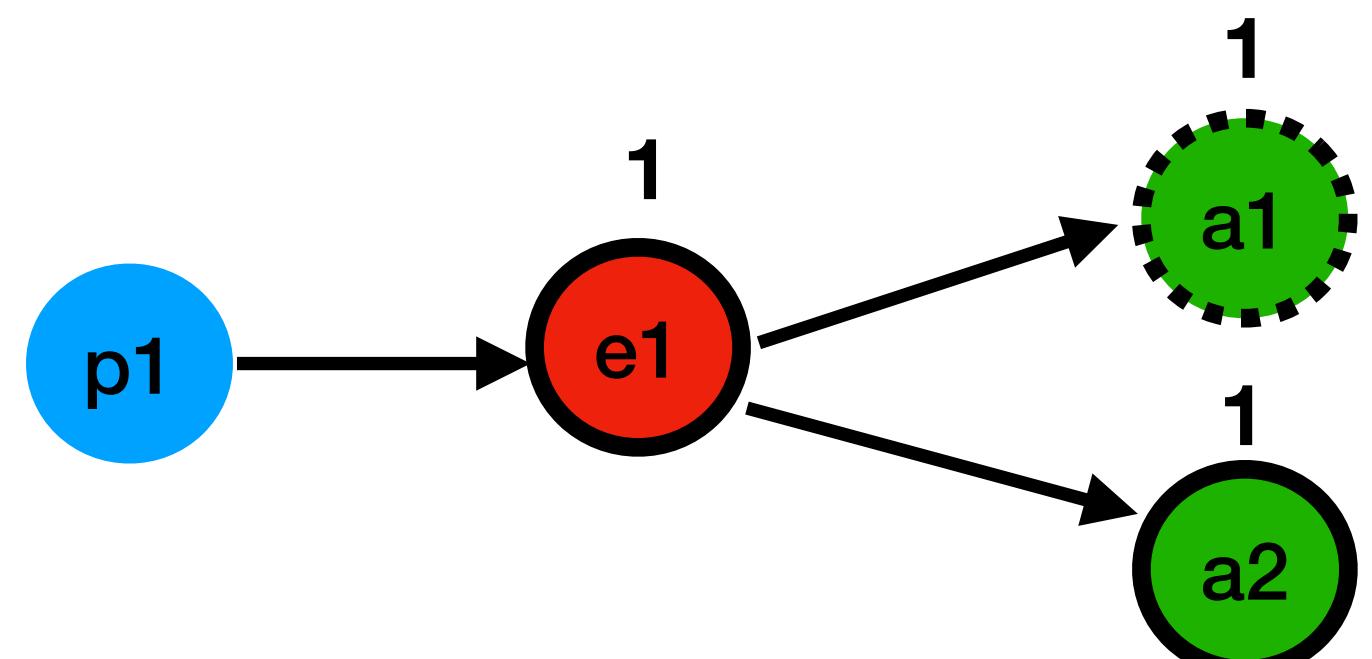
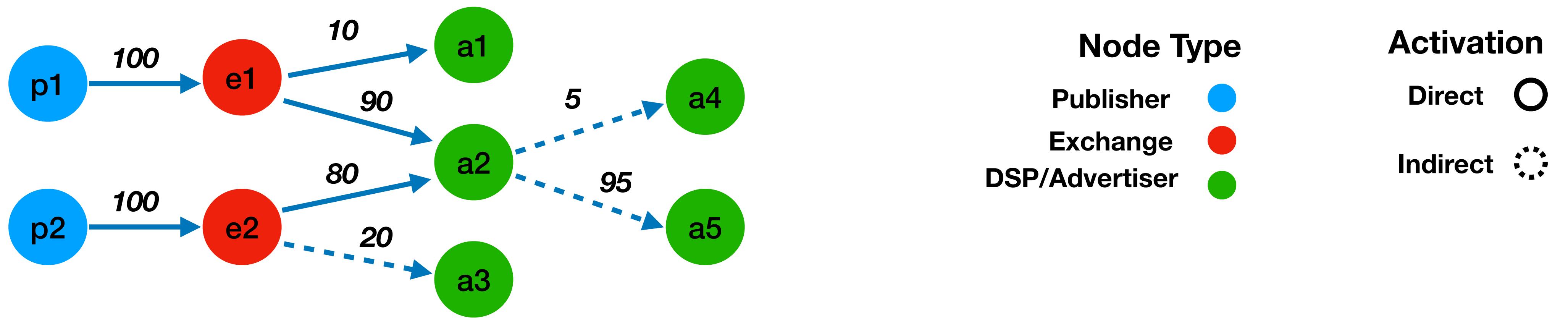
We manually select **37 exchanges**
which are allowed to forward
indirect impressions to solicit bids
during RTB

Simulation Example



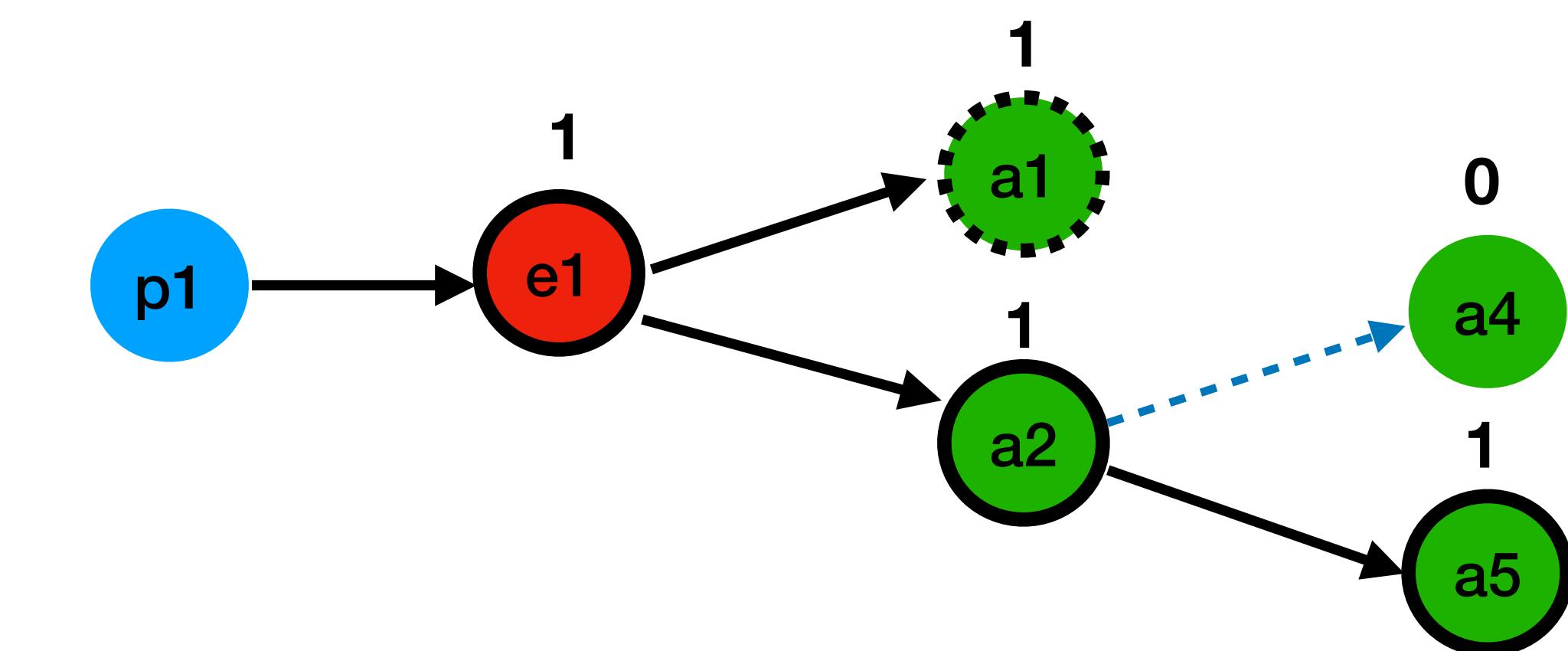
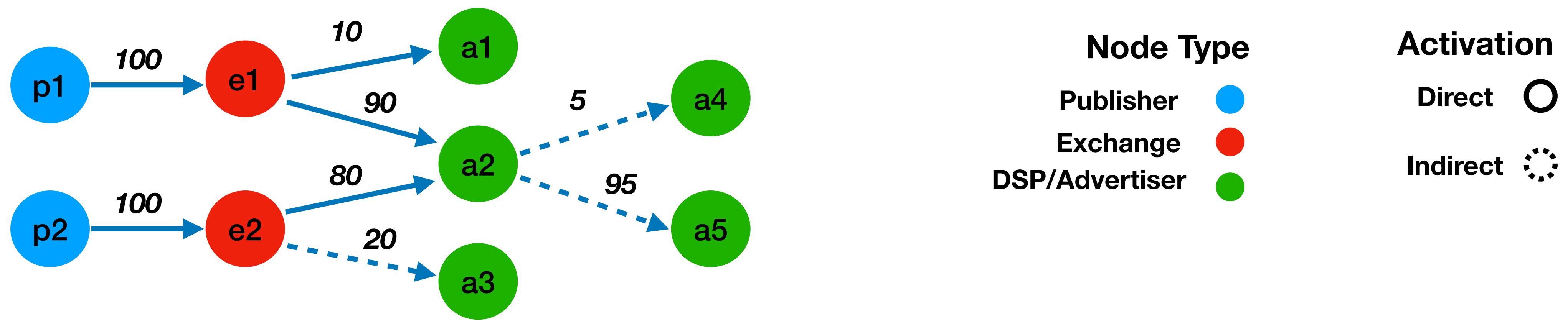
We manually select **37 exchanges** which are allowed to forward **indirect** impressions to solicit bids during RTB

Simulation Example



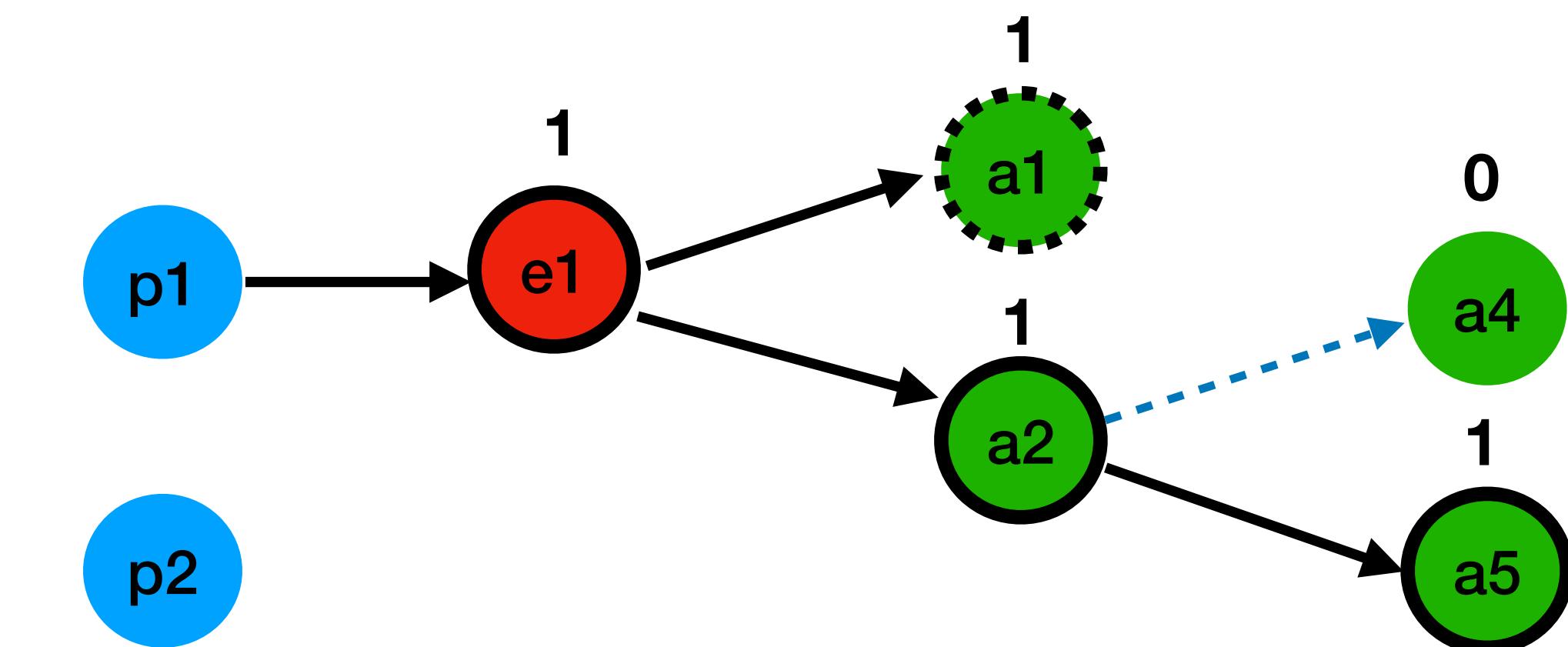
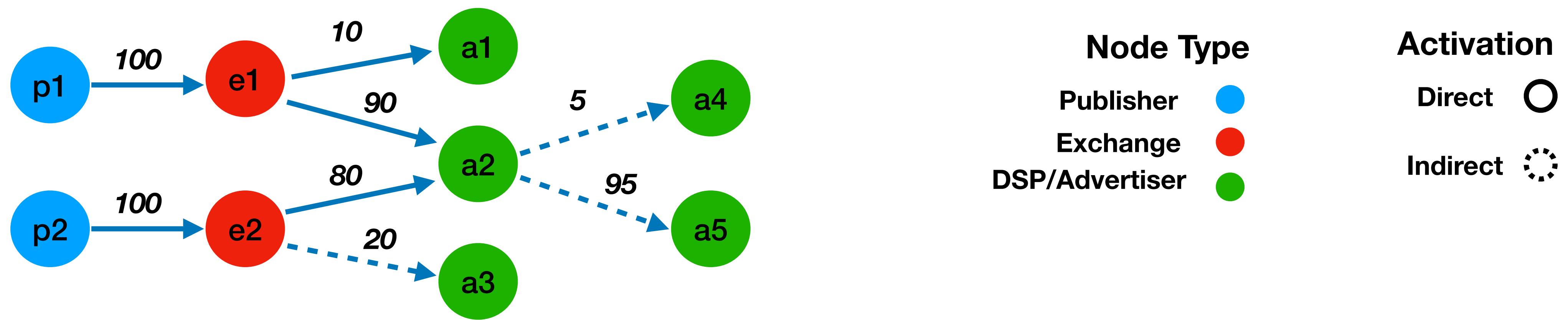
We manually select **37 exchanges** which are allowed to forward **indirect** impressions to solicit bids during RTB

Simulation Example



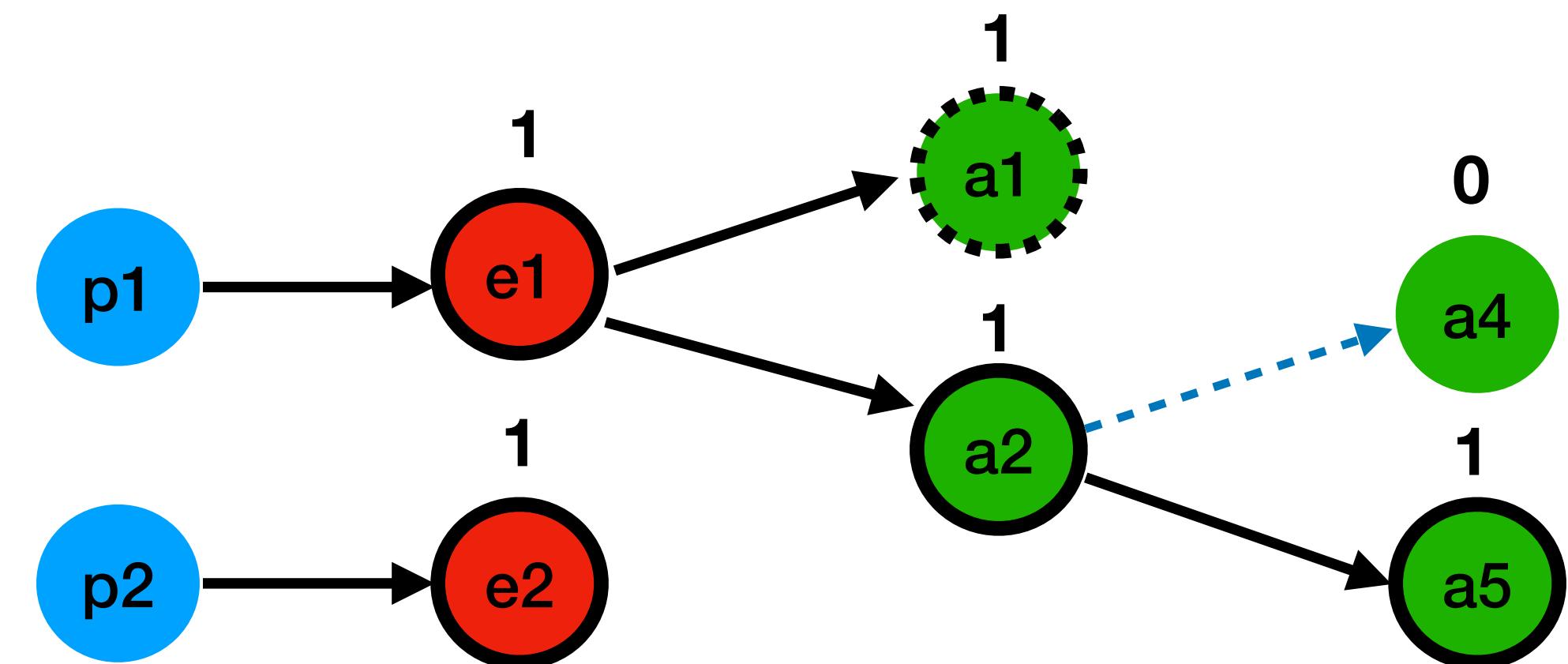
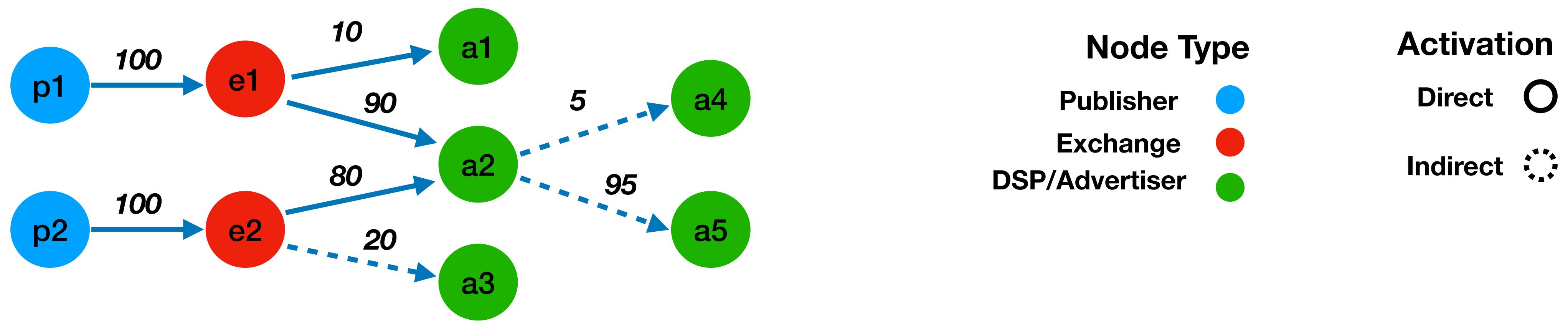
We manually select **37 exchanges** which are allowed to forward **indirect** impressions to solicit bids during RTB

Simulation Example



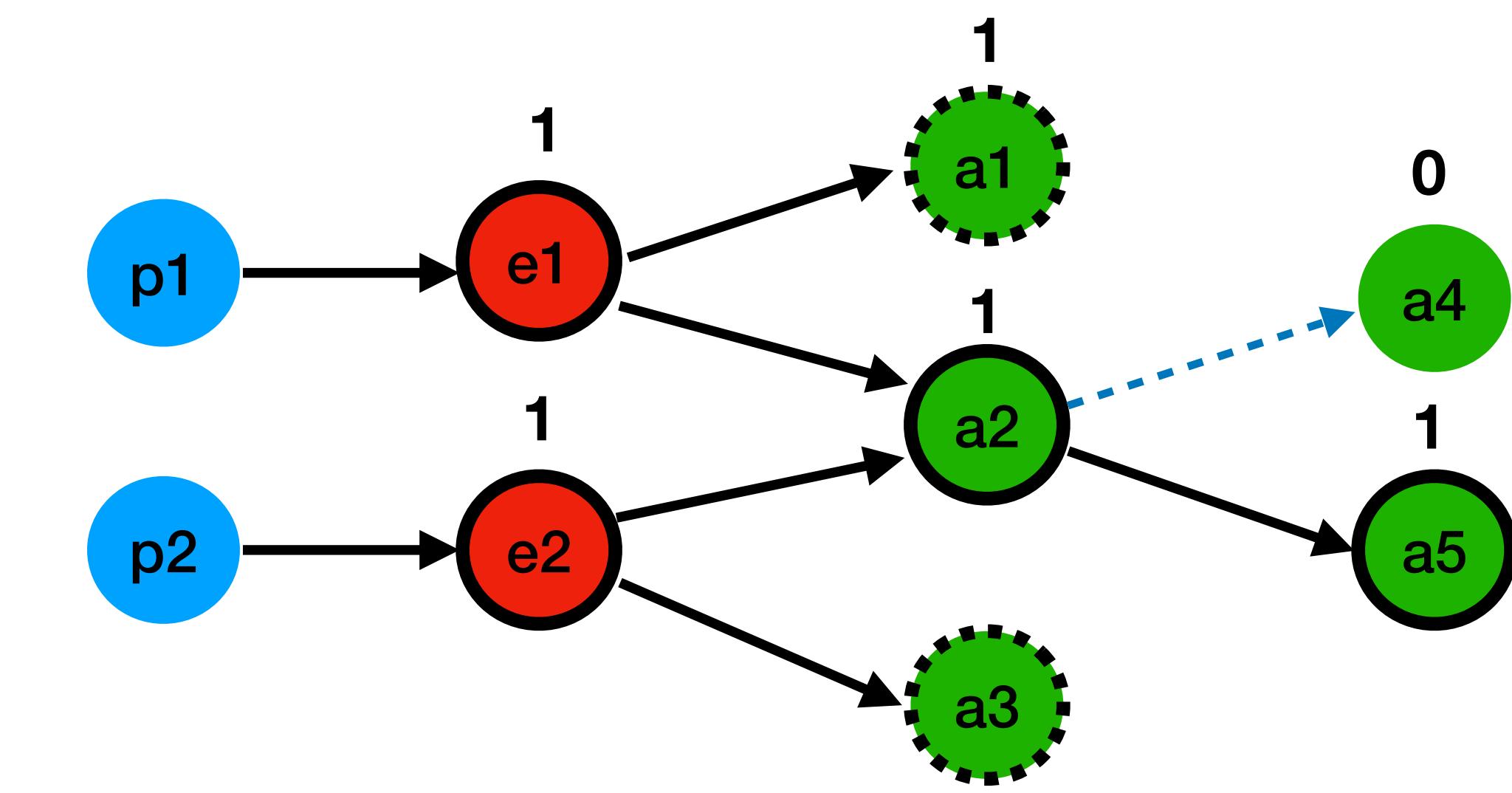
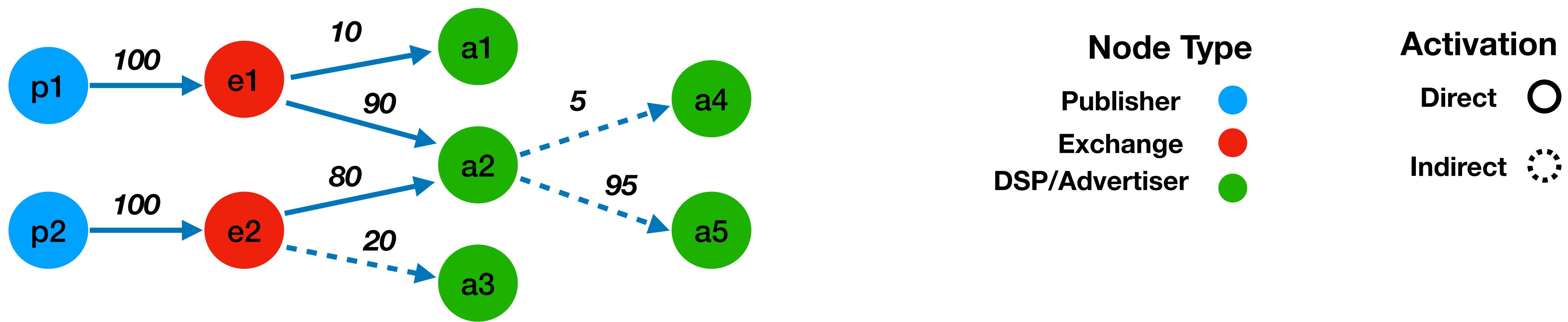
We manually select **37 exchanges** which are allowed to forward **indirect** impressions to solicit bids during RTB

Simulation Example



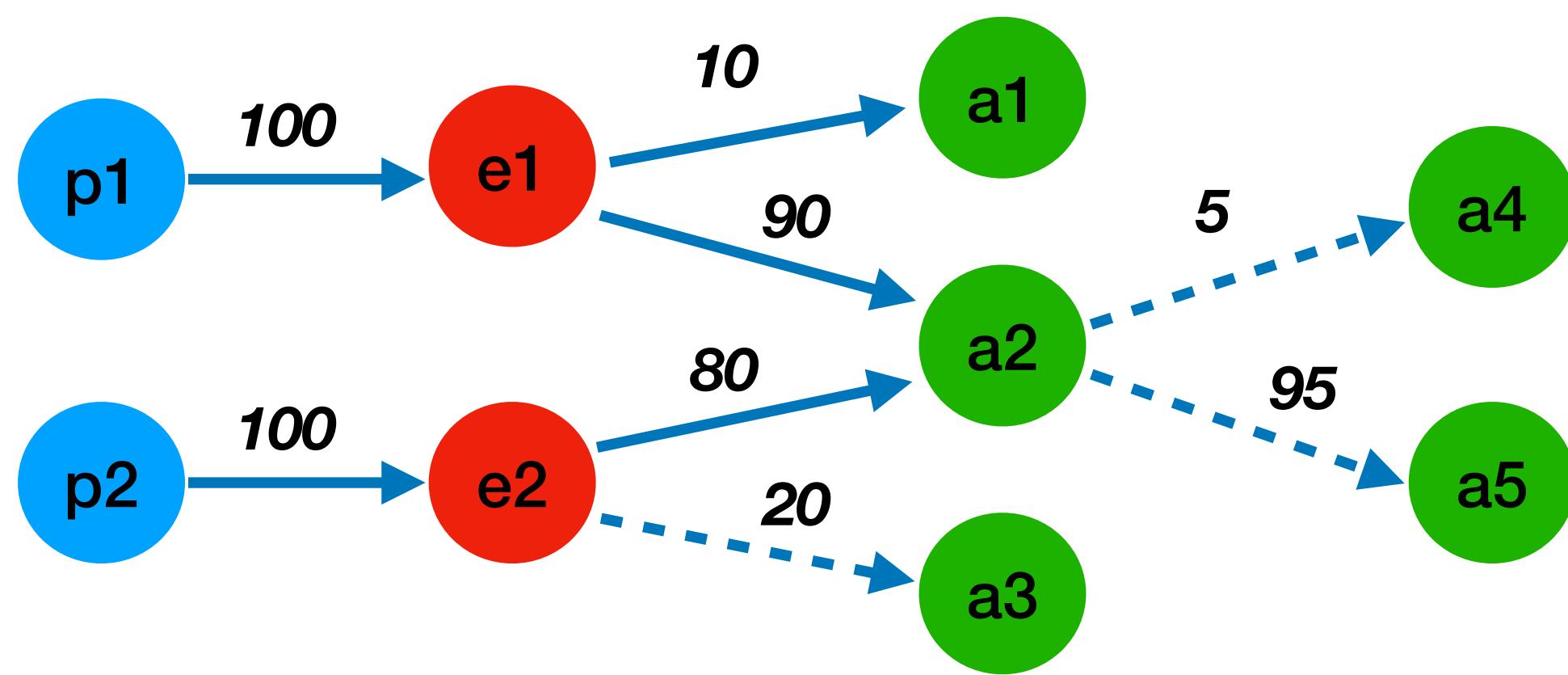
We manually select **37 exchanges** which are allowed to forward **indirect** impressions to solicit bids during RTB

Simulation Example

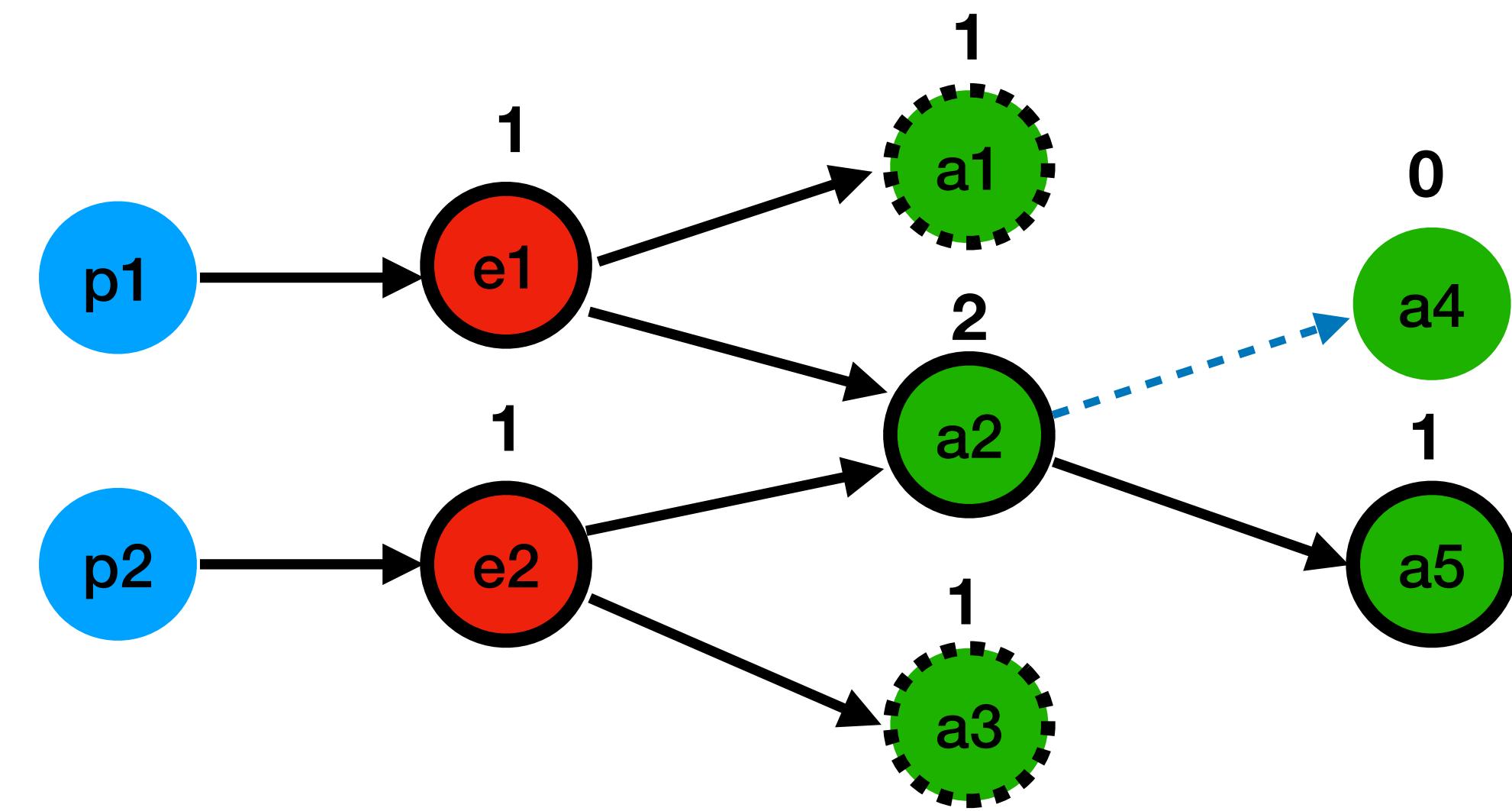


We manually select **37 exchanges** which are allowed to forward **indirect** impressions to solicit bids during RTB

Simulation Example

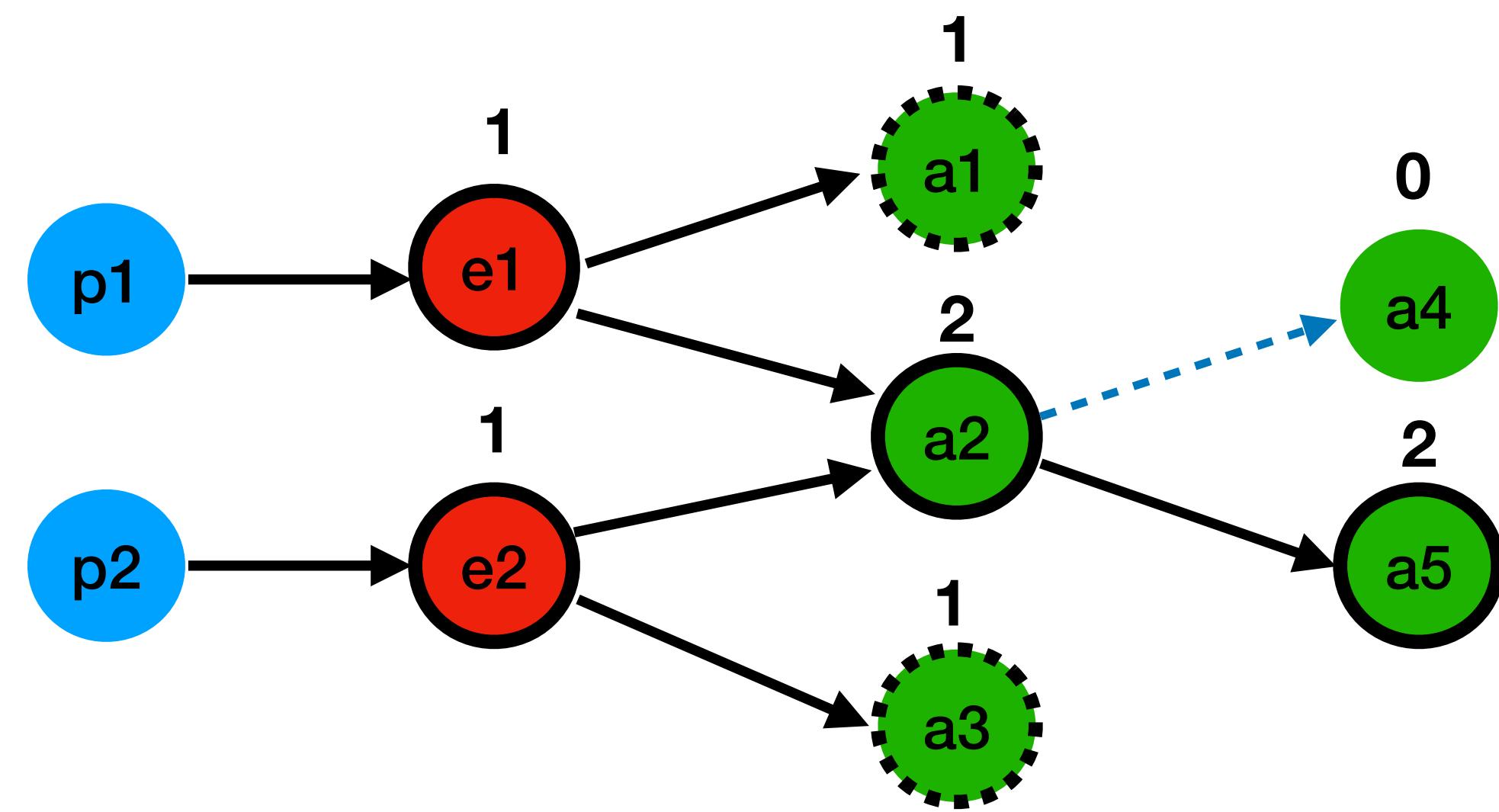
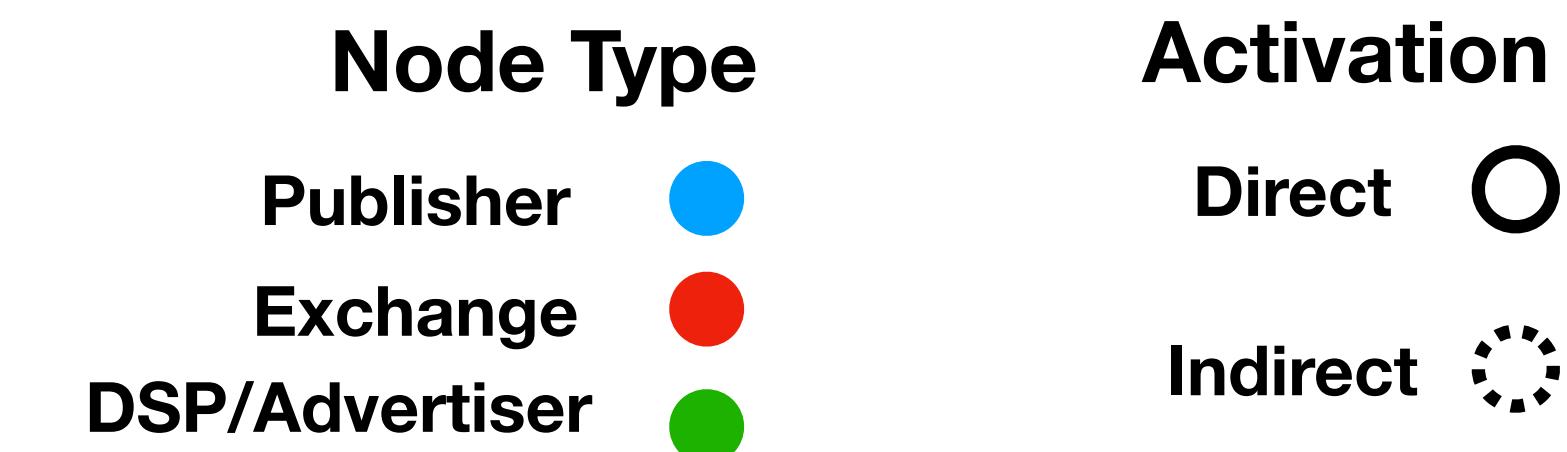
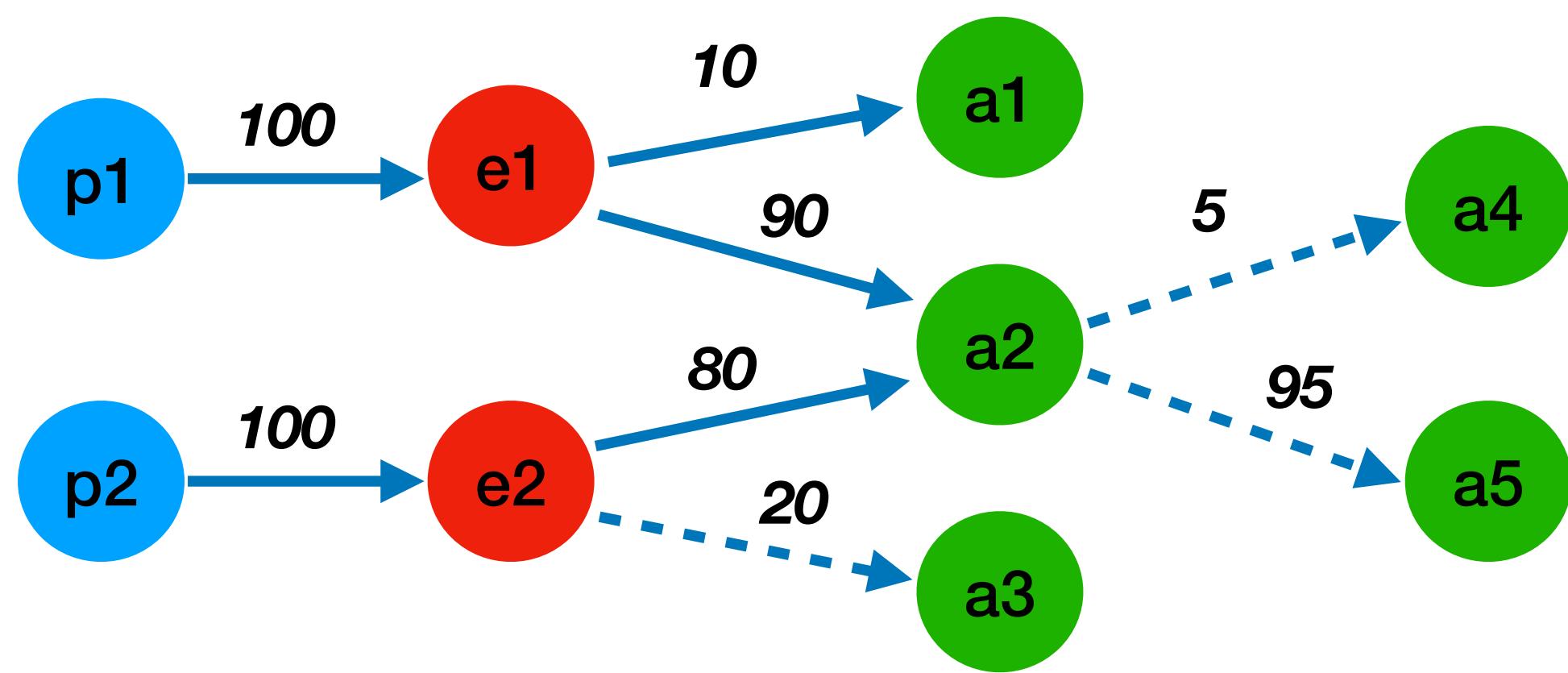


Node Type	Activation
Publisher	Direct
Exchange	Indirect
DSP/Advertiser	Indirect



We manually select **37 exchanges** which are allowed to forward **indirect** impressions to solicit bids during RTB

Simulation Example



We manually select **37 exchanges** which are allowed to forward **indirect** impressions to solicit bids during RTB

Impressions Observed

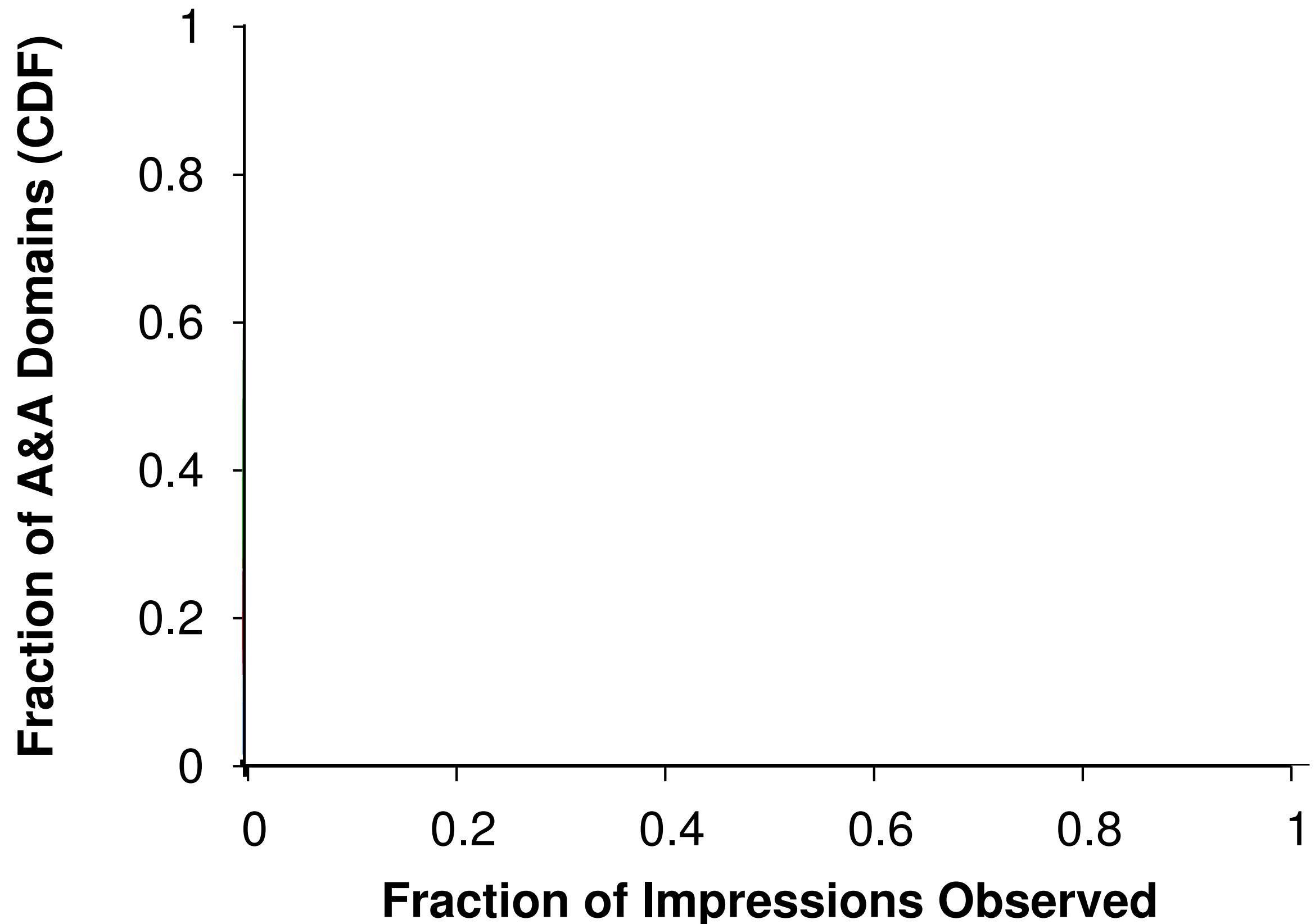
We have 3 simulation models:

1. **RTB-Relaxed**: Upper-bound
2. **Cookie-Matching**: Lower-bound
3. **RTB-Constrained**: Realistic Scenario

Impressions Observed

We have 3 simulation models:

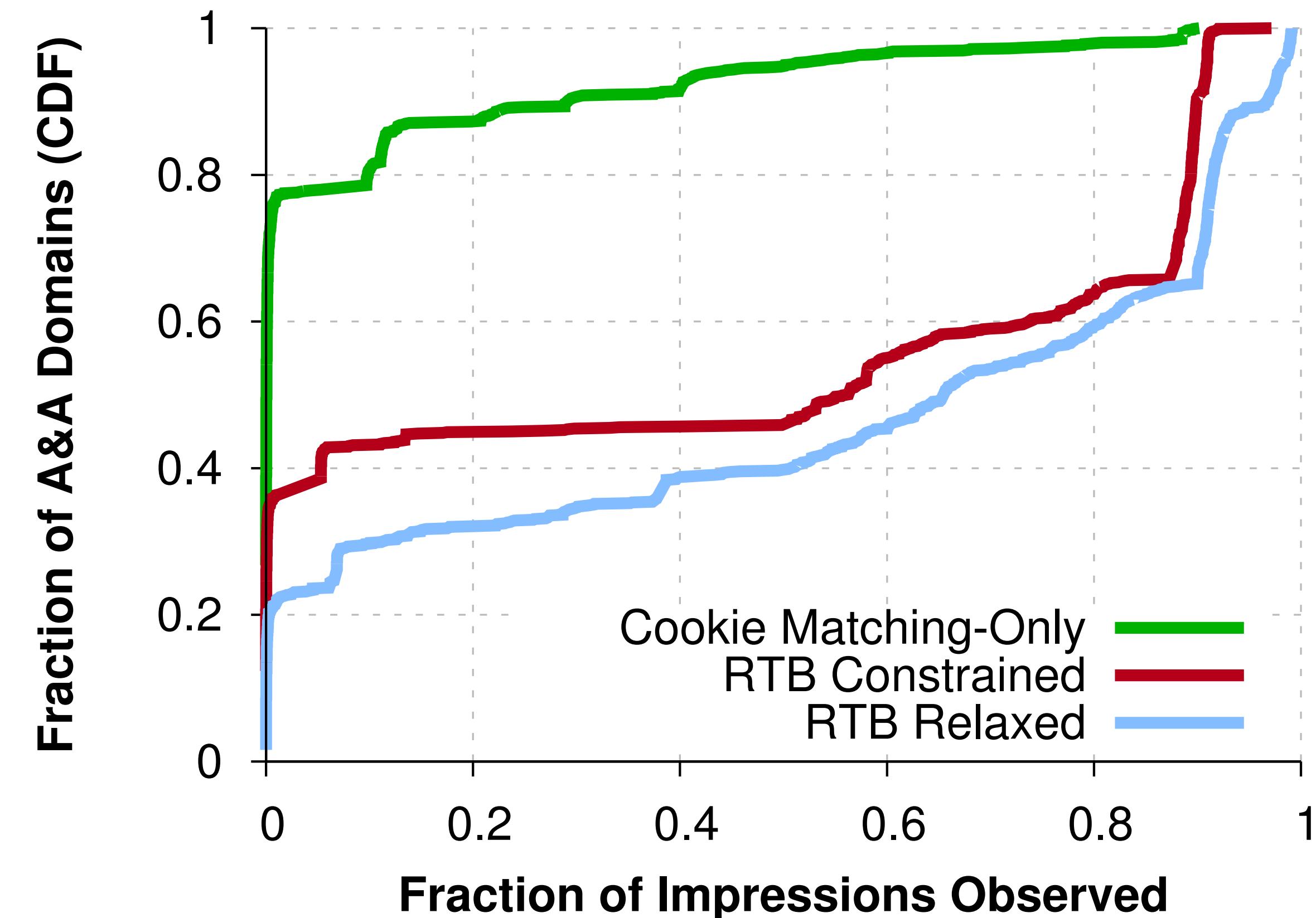
1. **RTB-Relaxed**: Upper-bound
2. **Cookie-Matching**: Lower-bound
3. **RTB-Constrained**: Realistic Scenario



Impressions Observed

We have 3 simulation models:

1. **RTB-Relaxed**: Upper-bound
2. **Cookie-Matching**: Lower-bound
3. **RTB-Constrained**: Realistic Scenario



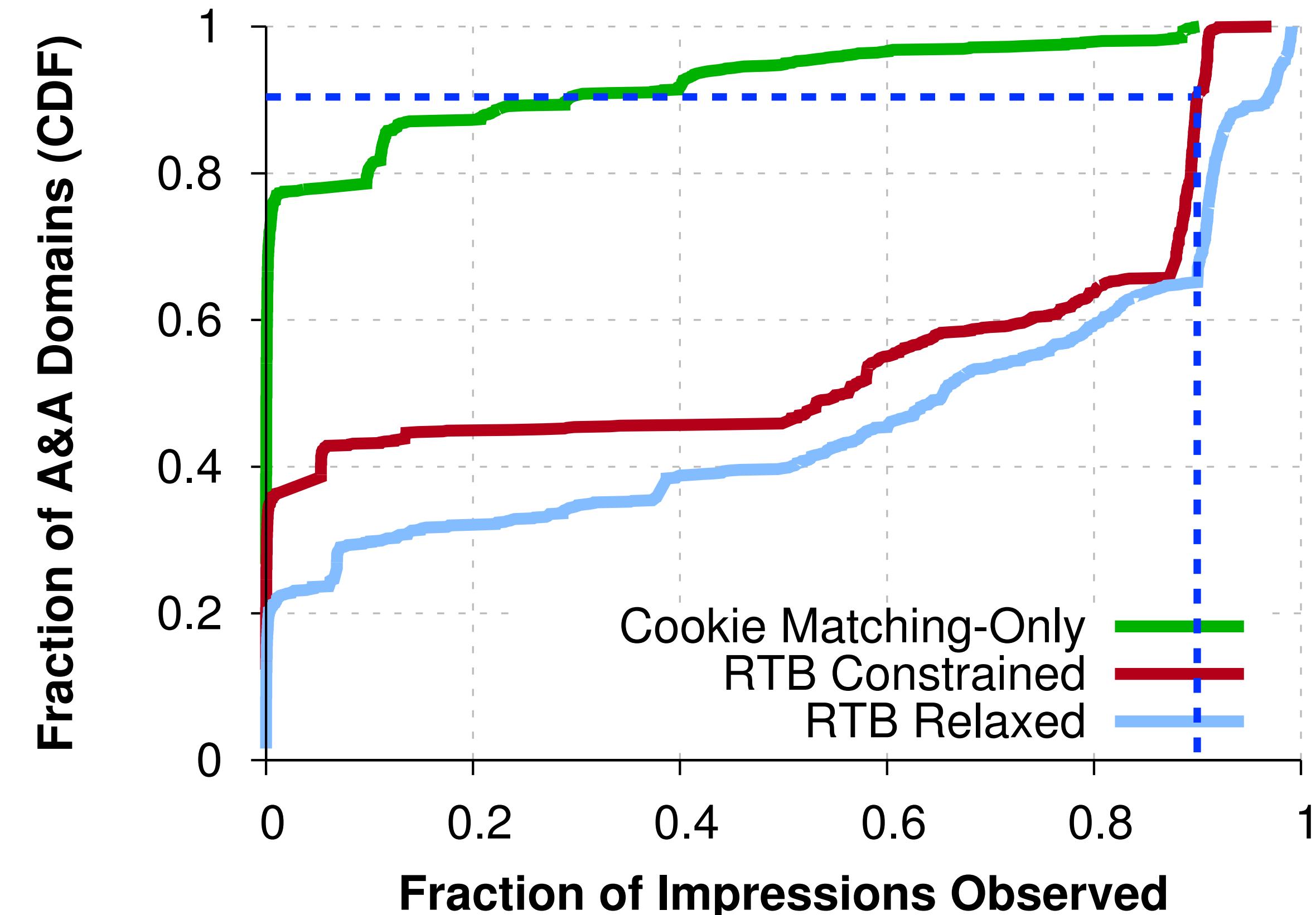
Impressions Observed

We have 3 simulation models:

1. **RTB-Relaxed:** Upper-bound
2. **Cookie-Matching:** Lower-bound
3. **RTB-Constrained:** Realistic Scenario

Take Away

1. RTB-Constrained is very close to RTB-Relaxed
2. 10% A&A see more than 90% of impressions in RTB-Constrained



Effect of Blocking Extensions

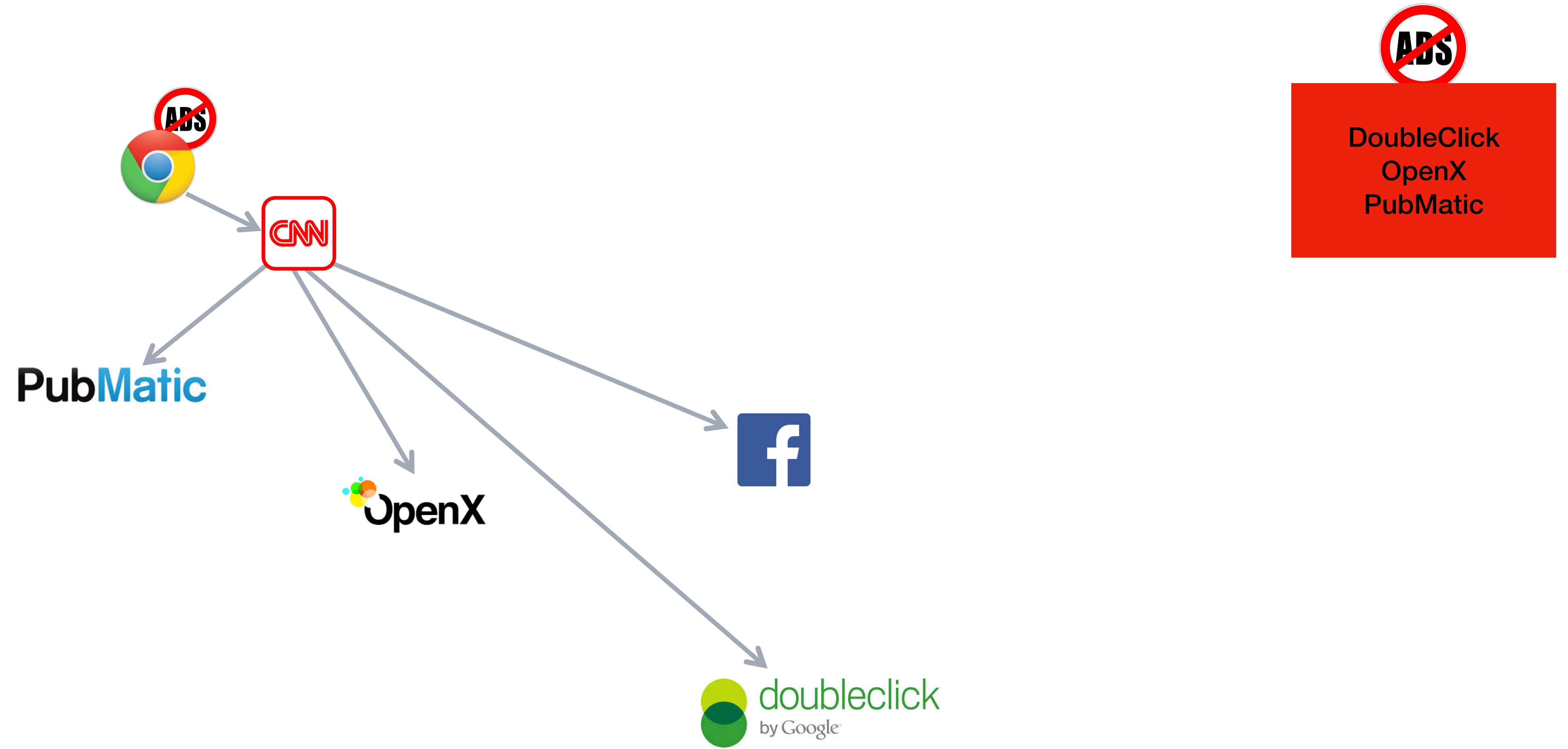


DoubleClick
OpenX
PubMatic

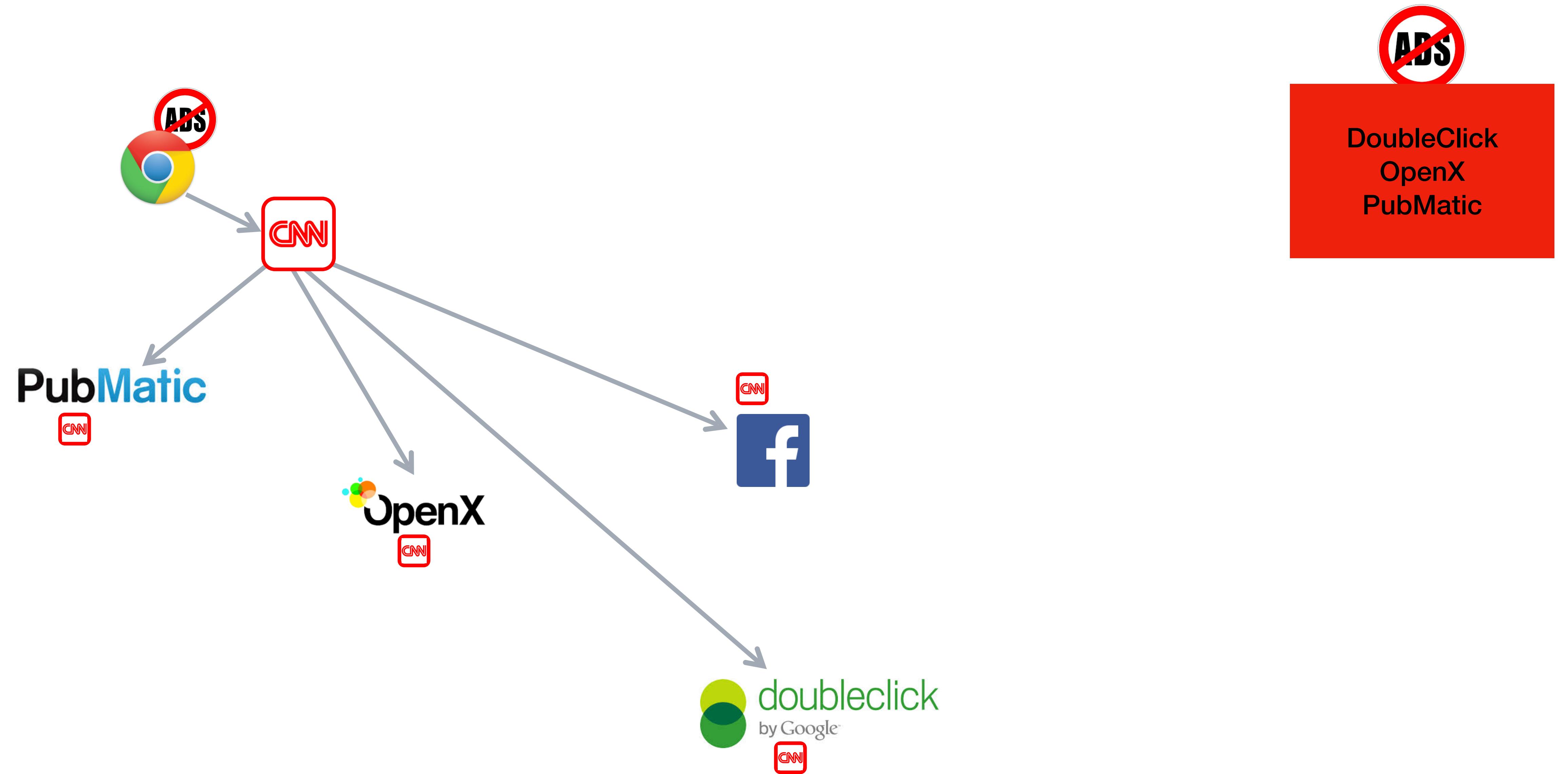
Effect of Blocking Extensions



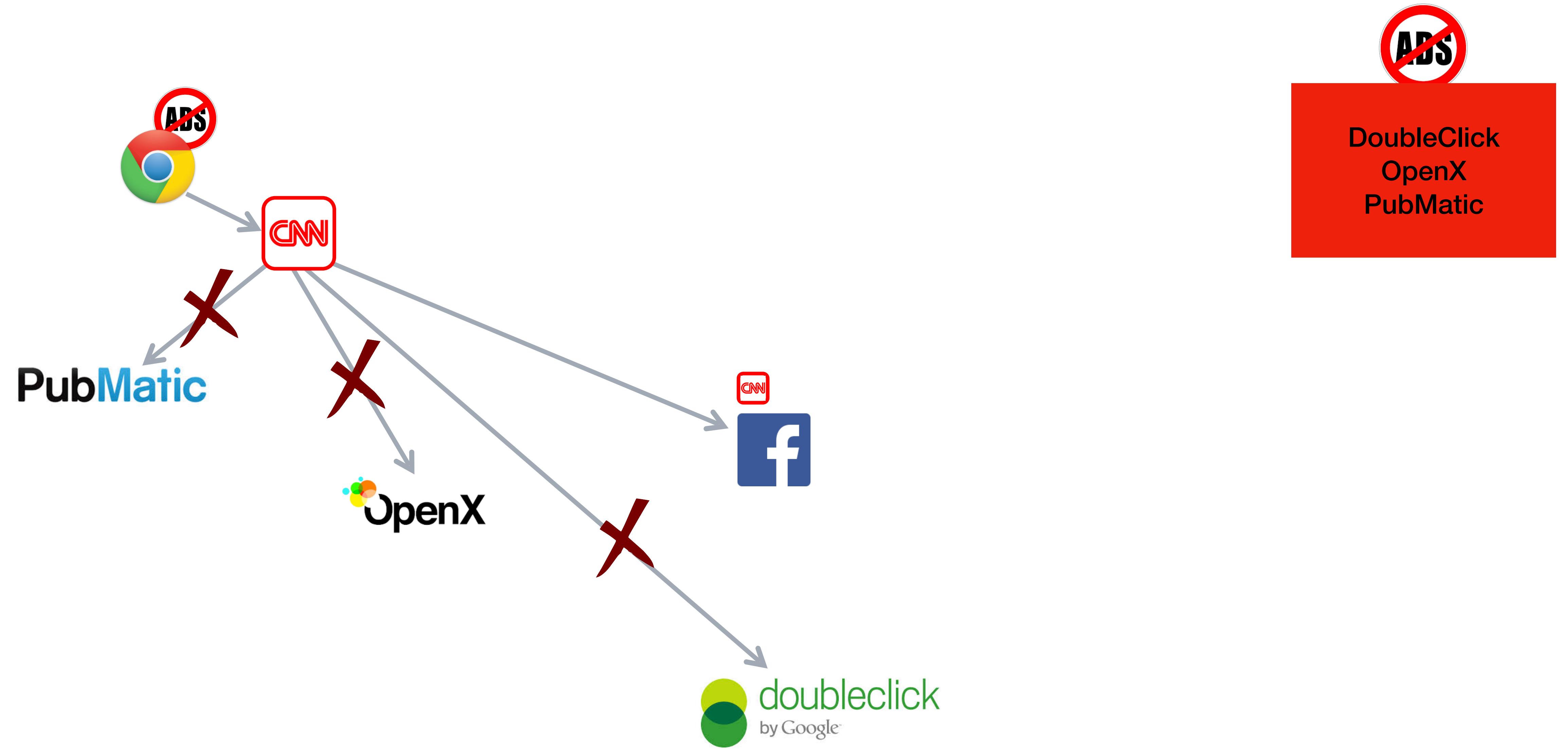
Effect of Blocking Extensions



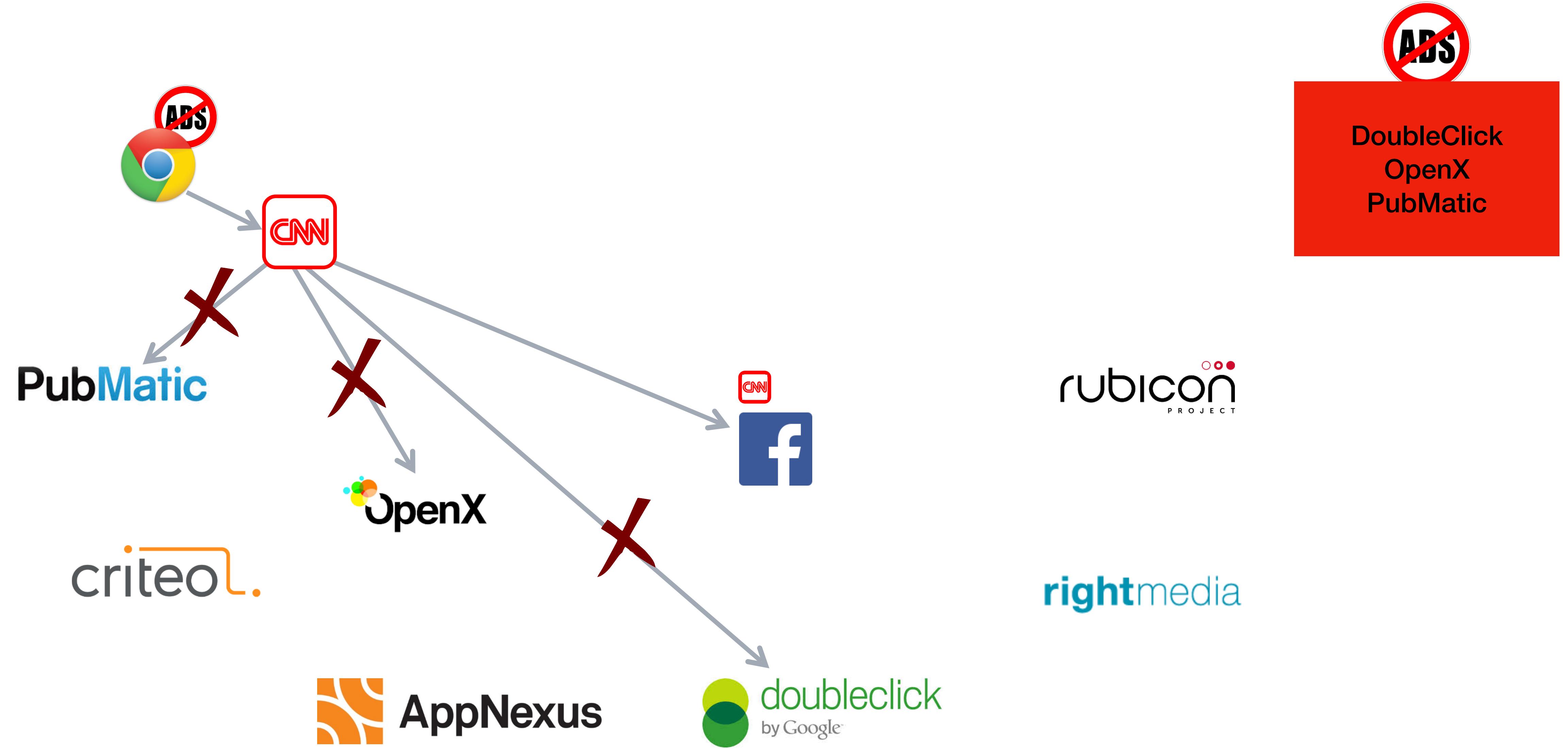
Effect of Blocking Extensions



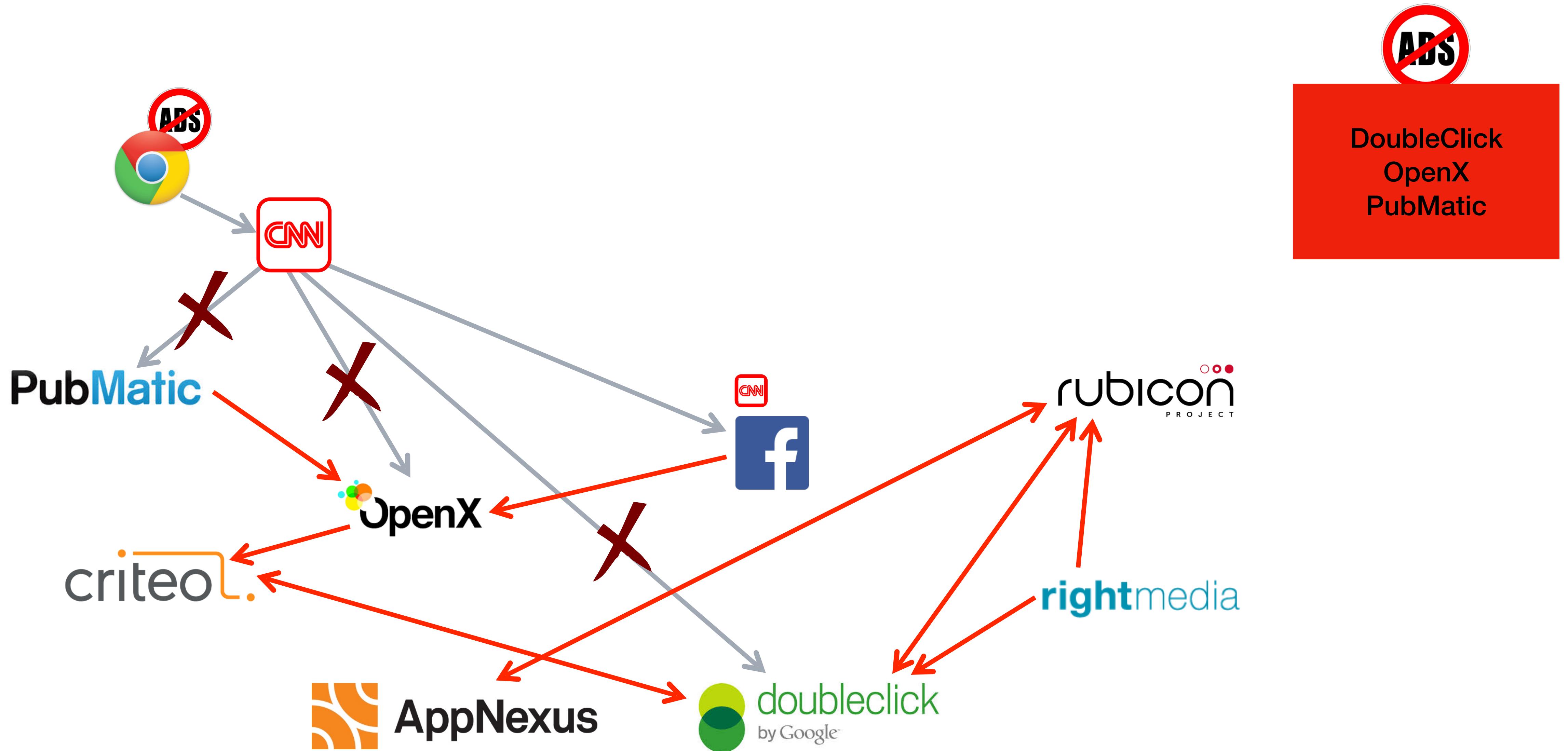
Effect of Blocking Extensions



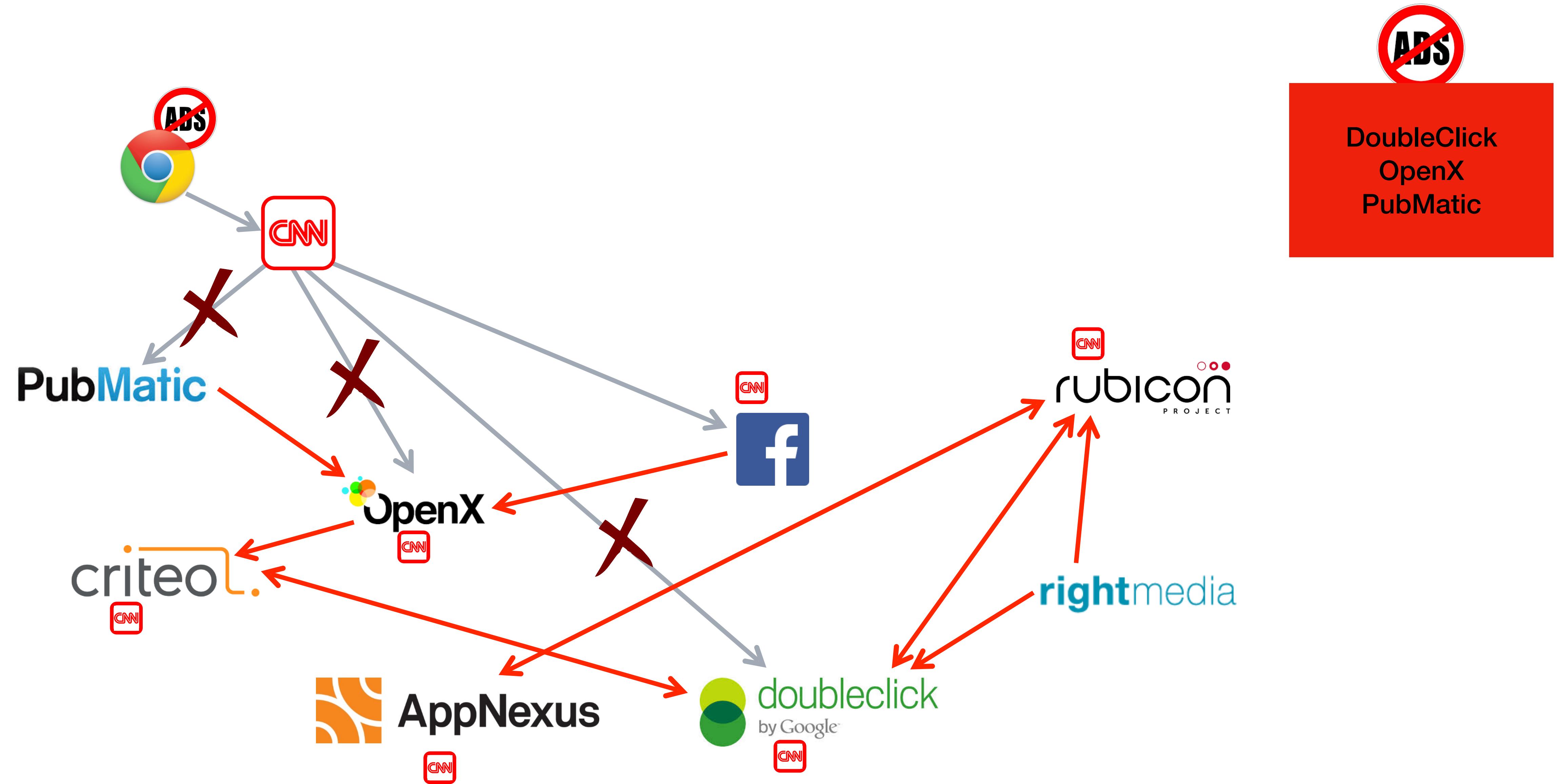
Effect of Blocking Extensions



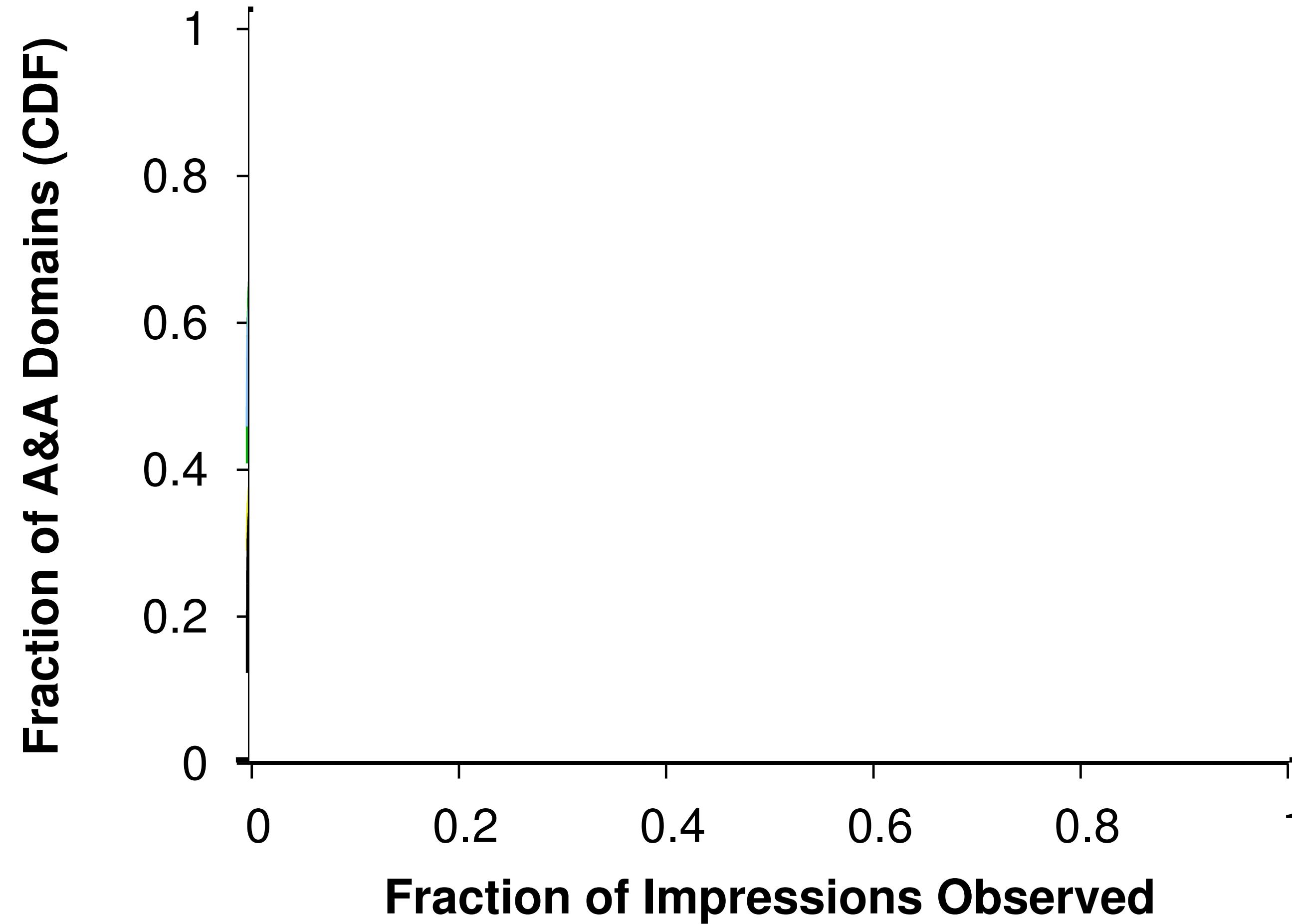
Effect of Blocking Extensions



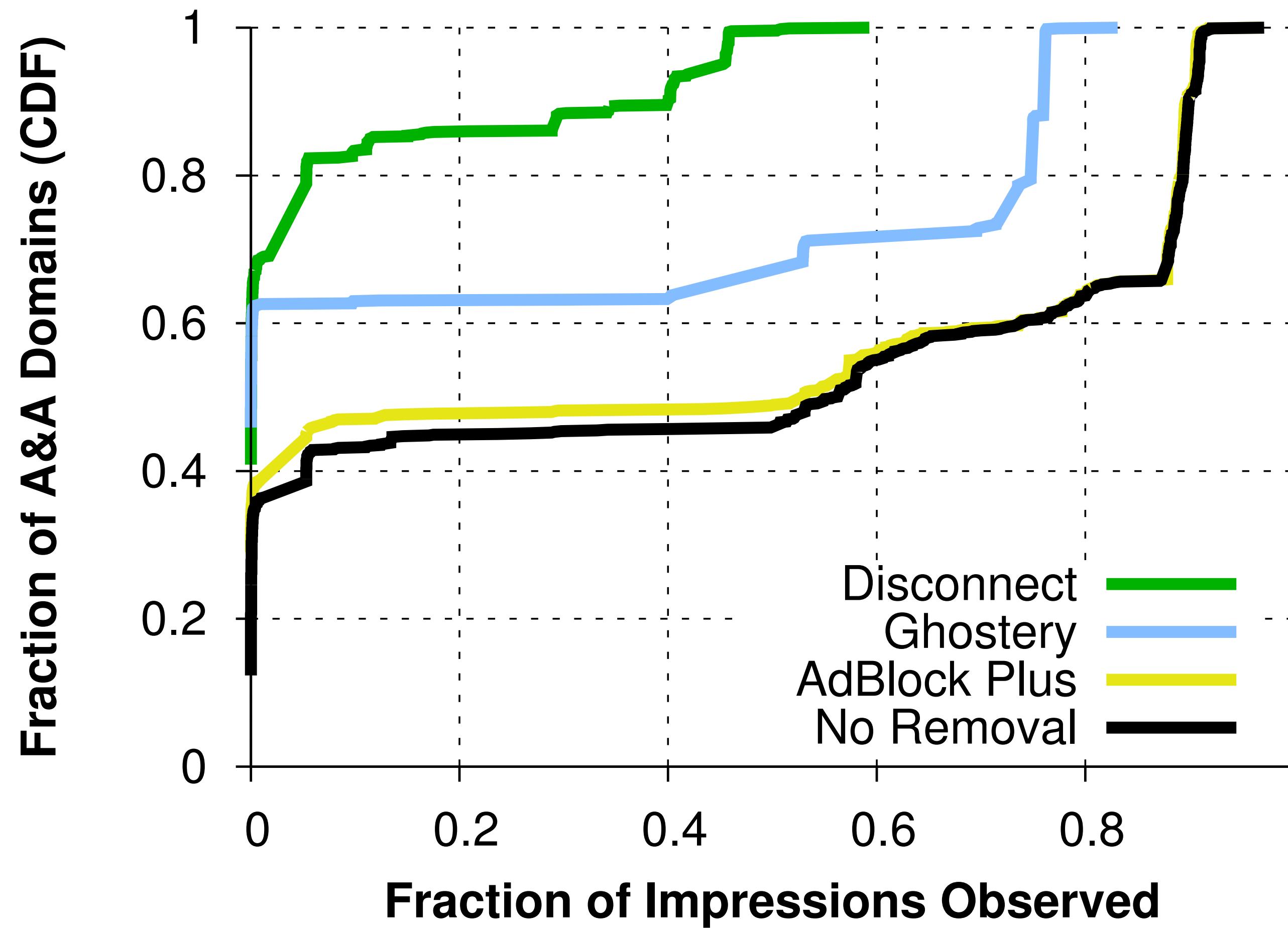
Effect of Blocking Extensions



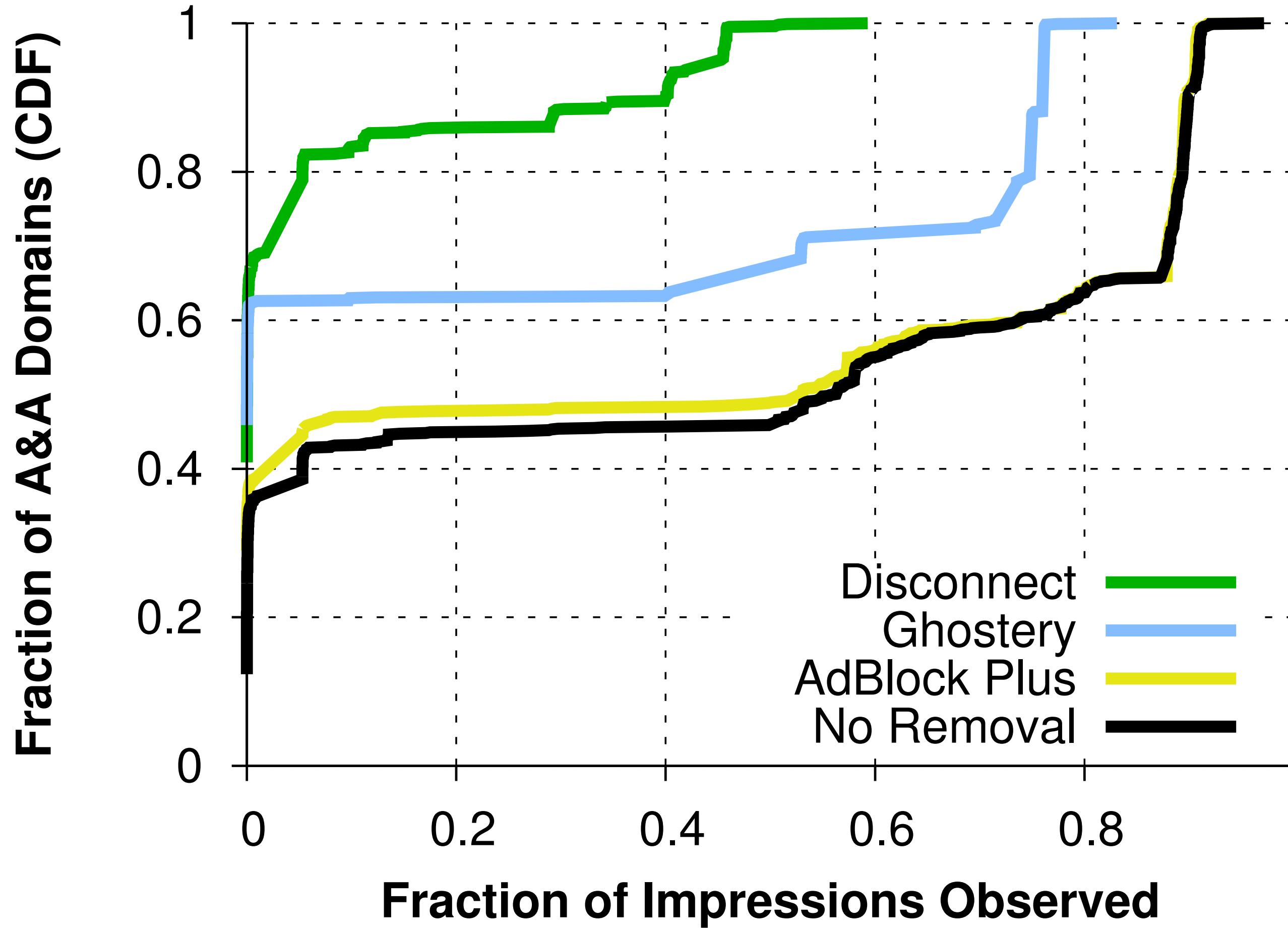
Impressions Observed With Blocking



Impressions Observed With Blocking



Impressions Observed With Blocking

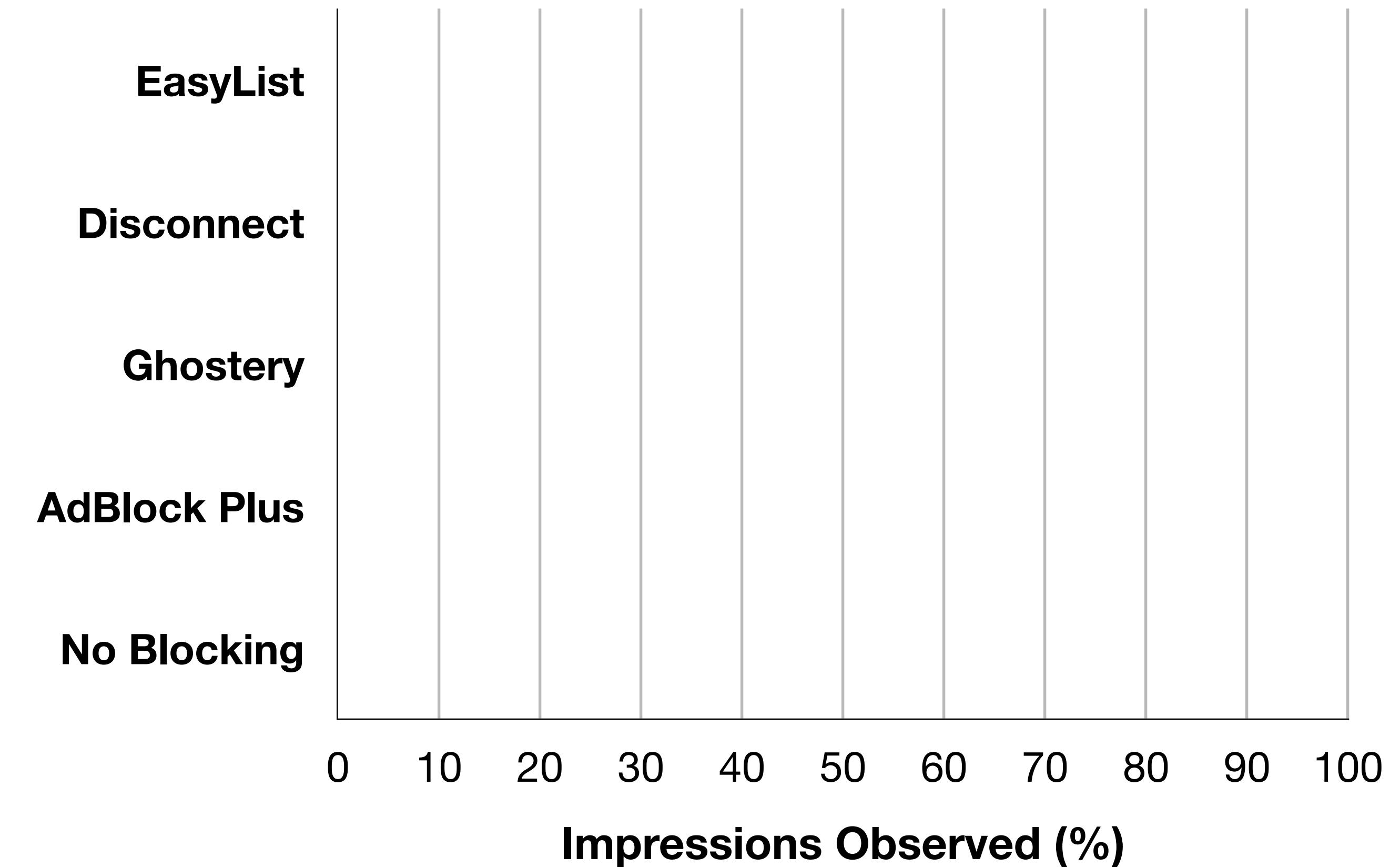


Take Away

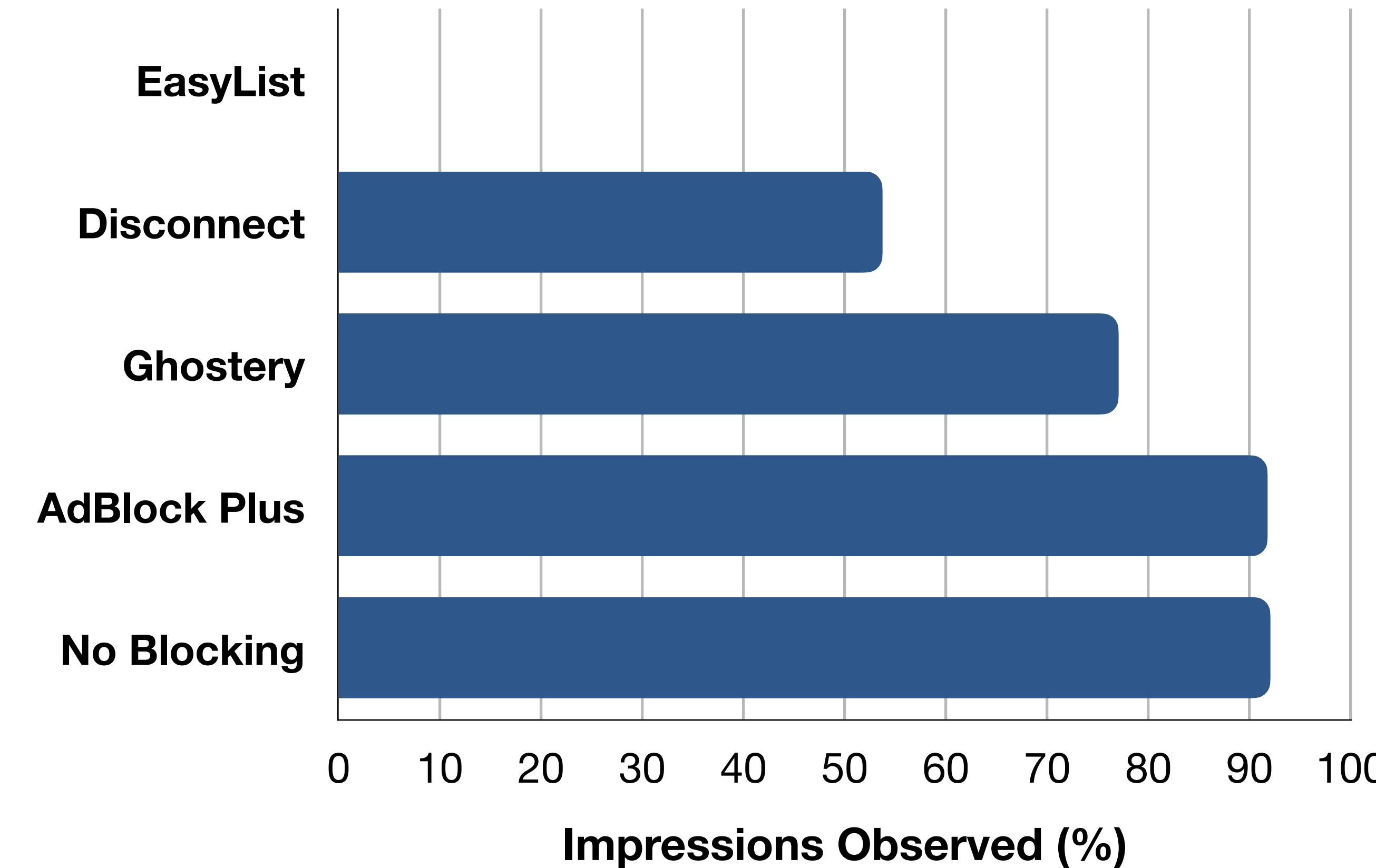
- Disconnect list is most effective.
- ABP is not effective at all due to Acceptable Ads program.
- Due to RTB, impressions are leaked to A&A domains even with blocking extensions.

Top 10 Domains Observing Impressions

Top 10 Domains Observing Impressions

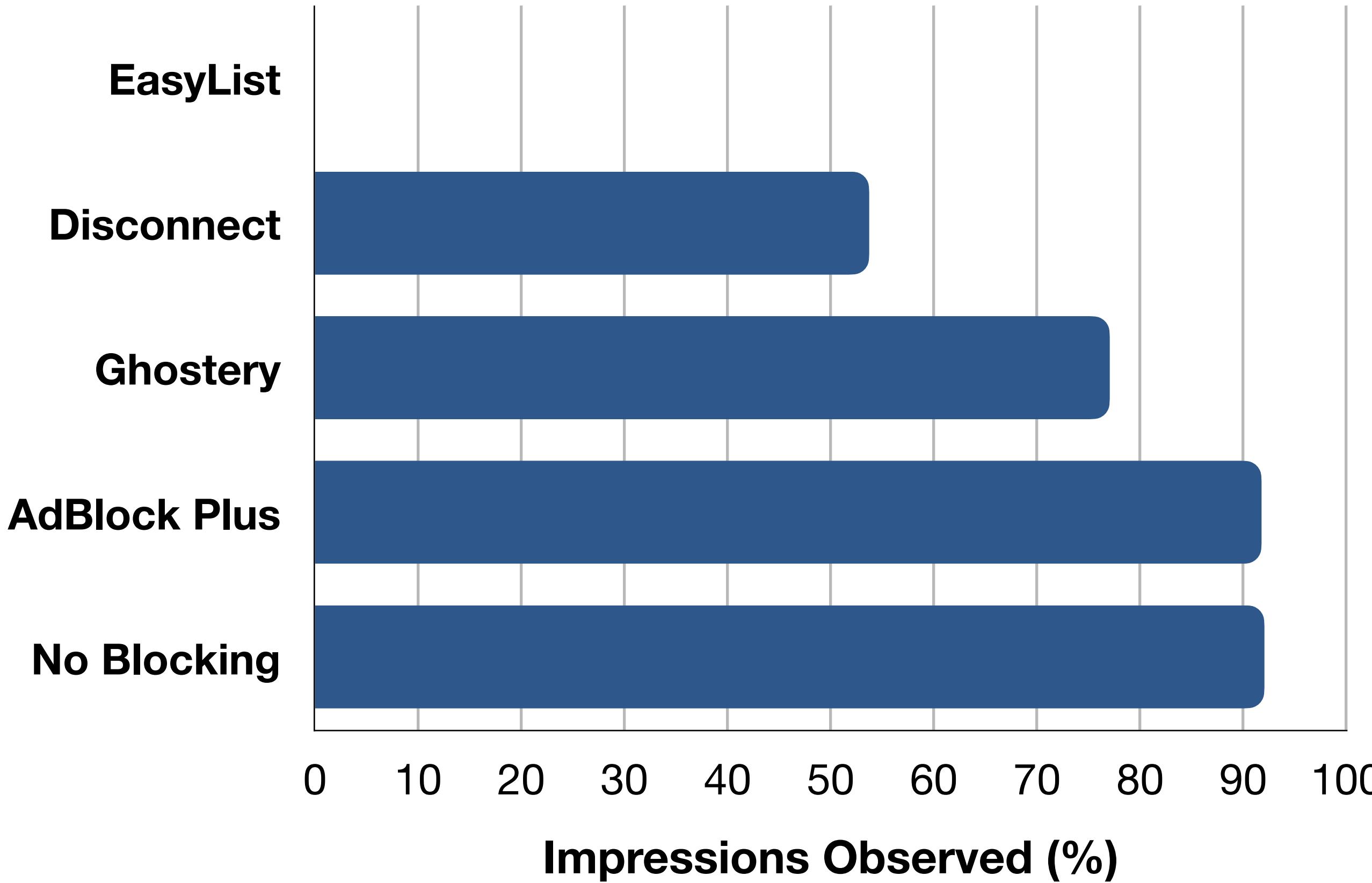


Top 10 Domains Observing Impressions



Top 10 domains can view majority of user impressions even with (most) blocking extensions installed

Top 10 Domains Observing Impressions



Domain	Impression %
google-analytics	97.0
youtube	91.7
quantserve	91.6
scorecardresearch	91.6
skimresources	91.3
twitter	91.1
pinterest	91.0
addthis	90.0
criteo	90.0
bluekai	90.8

Top 10 domains with most observed impressions under AdBlock Plus

Top 10 domains can view majority of user impressions even with (most) blocking extensions installed

Limitations of Our Simulations

- Our simulation models provide approximations
- Different users might have different browsing behaviors
 - We only simulate with respect to popular publishers
- The ecosystem could have changed from when the dataset was collected (December 2015)
- Not representative of mobile advertising ecosystem

Key Takeaways from this Work

- We provide a model to study the impact of Real Time Bidding (RTB) on user privacy.
- Ad Exchanges share user impressions to facilitate RTB
 - More than 10% A&A domains view up to 90% of user impressions under realistic conditions.
- Due to RTB, impressions can leak to A&A domains even with blocking extensions
 - AdBlock Plus is not effective at all due to Acceptable Ads program
 - Disconnect performed the best in terms of protecting privacy

Outline

Background & Motivation

Limitations of Prior Work

Detecting Information Sharing

- Using Retargeted Ads

Propagation of Tracking Data

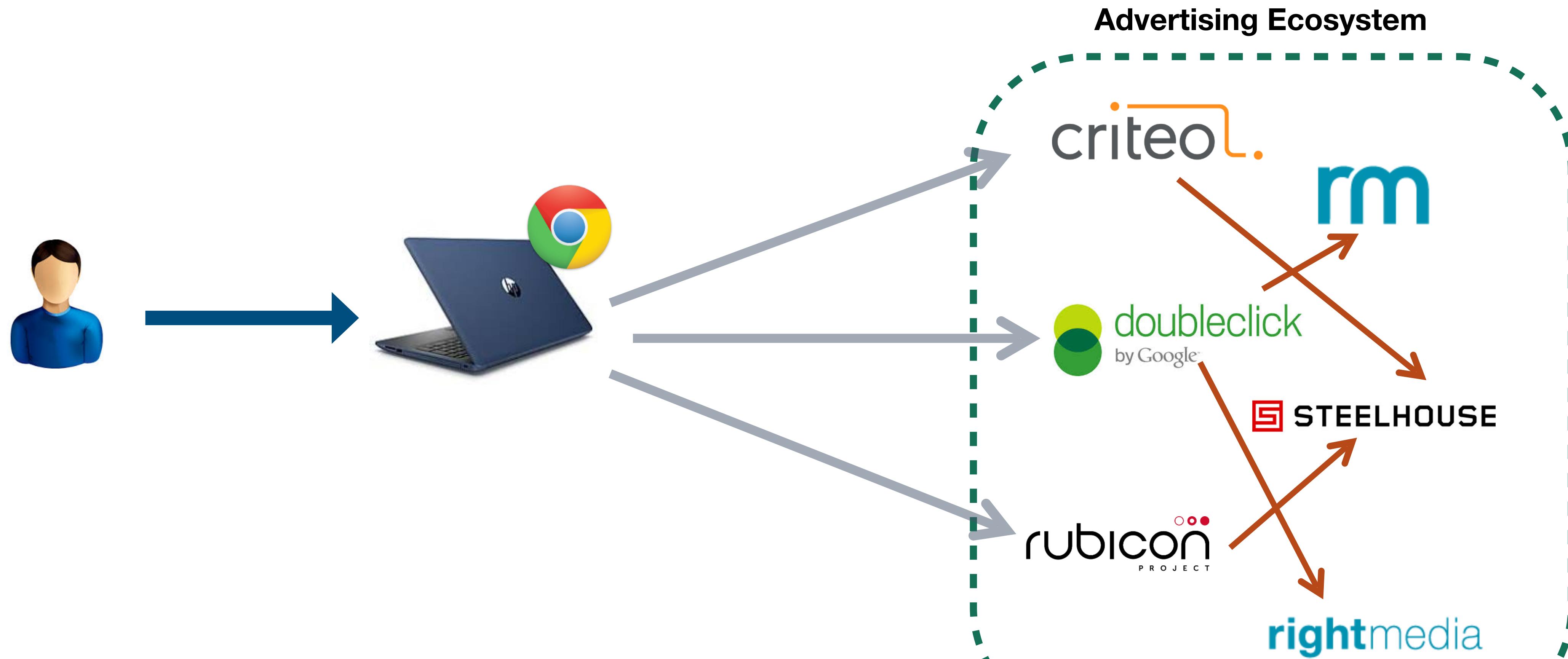
- Modeling via Graph

Ongoing Work

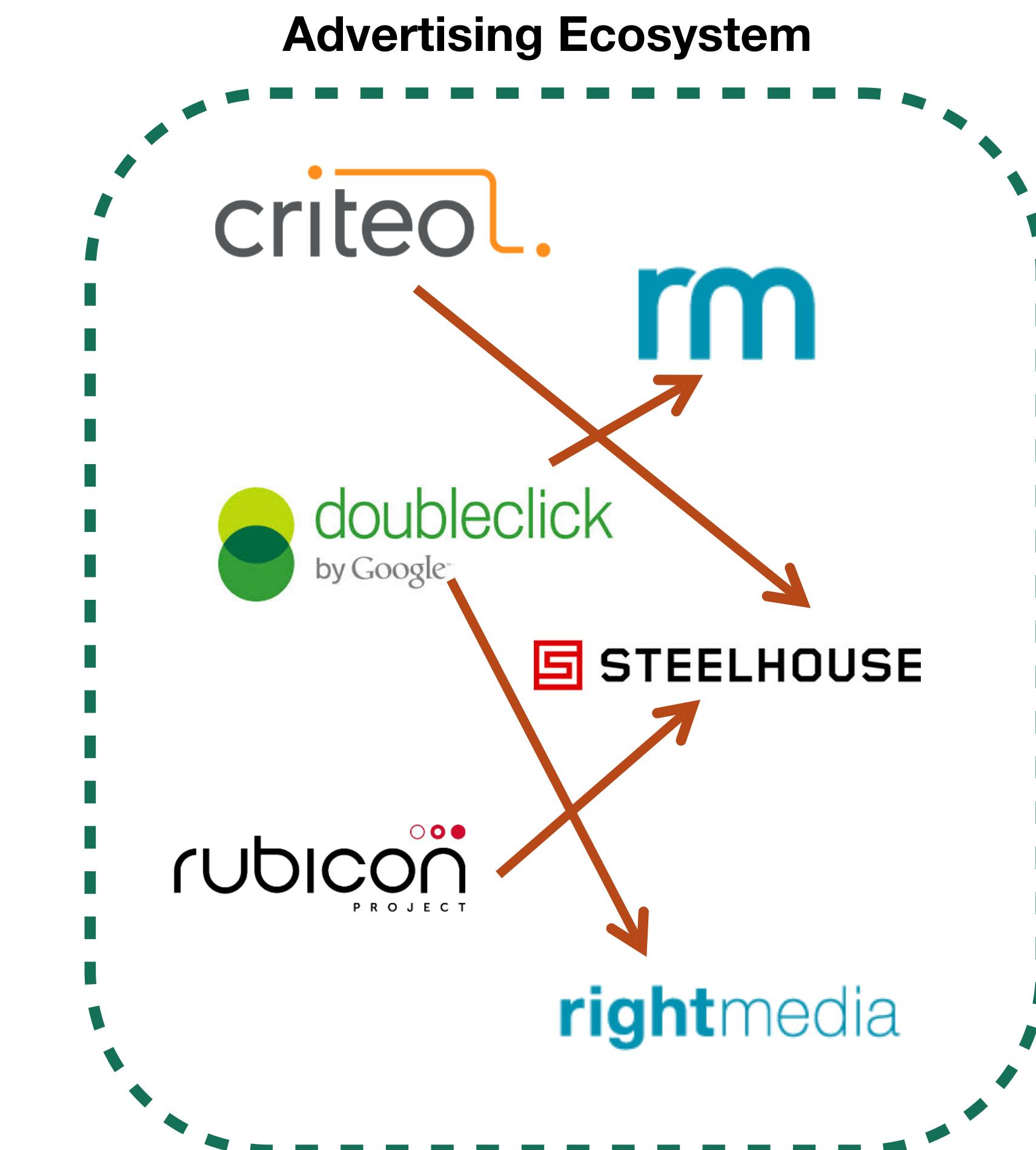
- Cross-device Tracking
- Automation of Adblocking Filter Lists

Research Plan

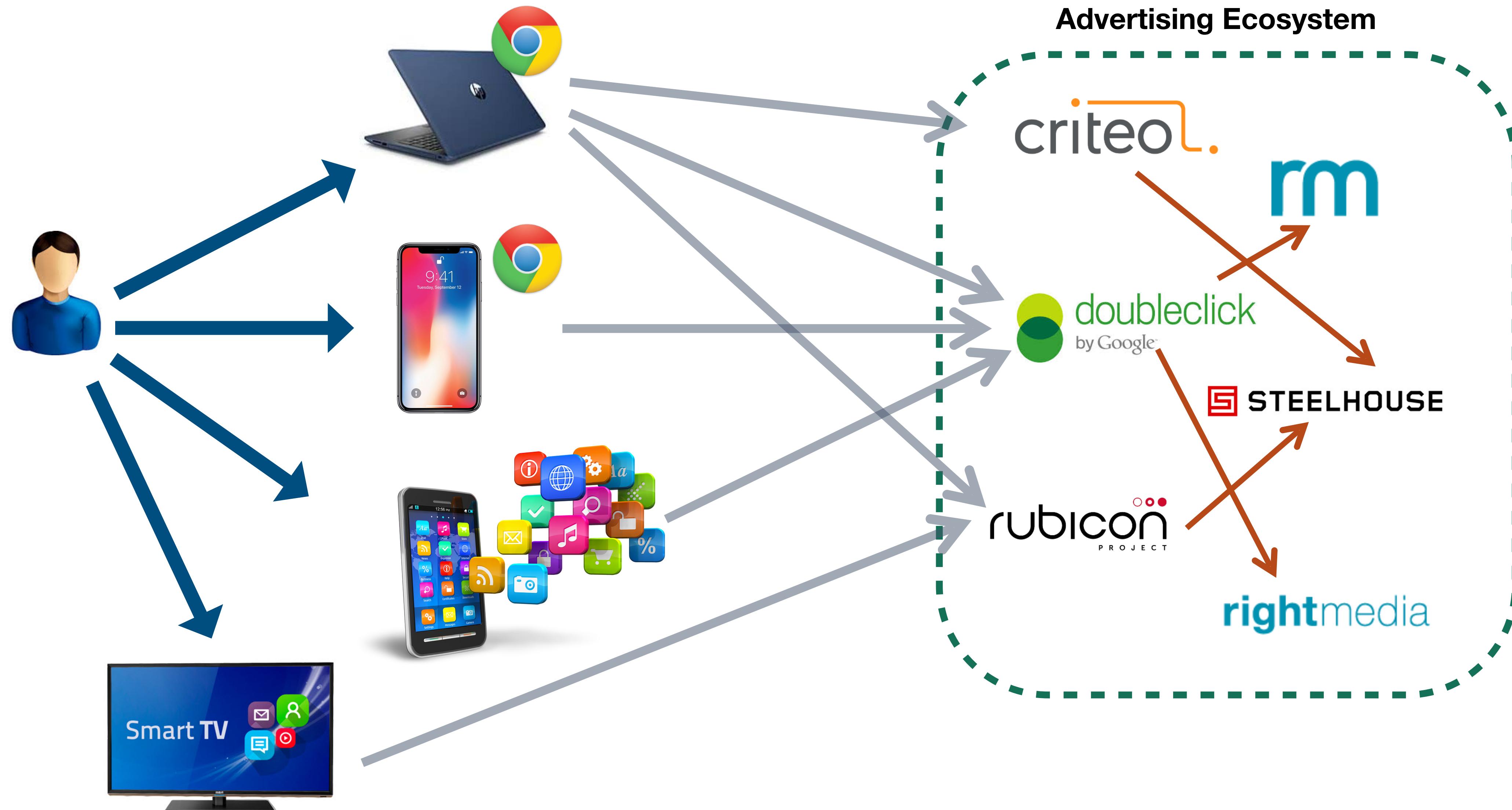
Cross Device Tracking



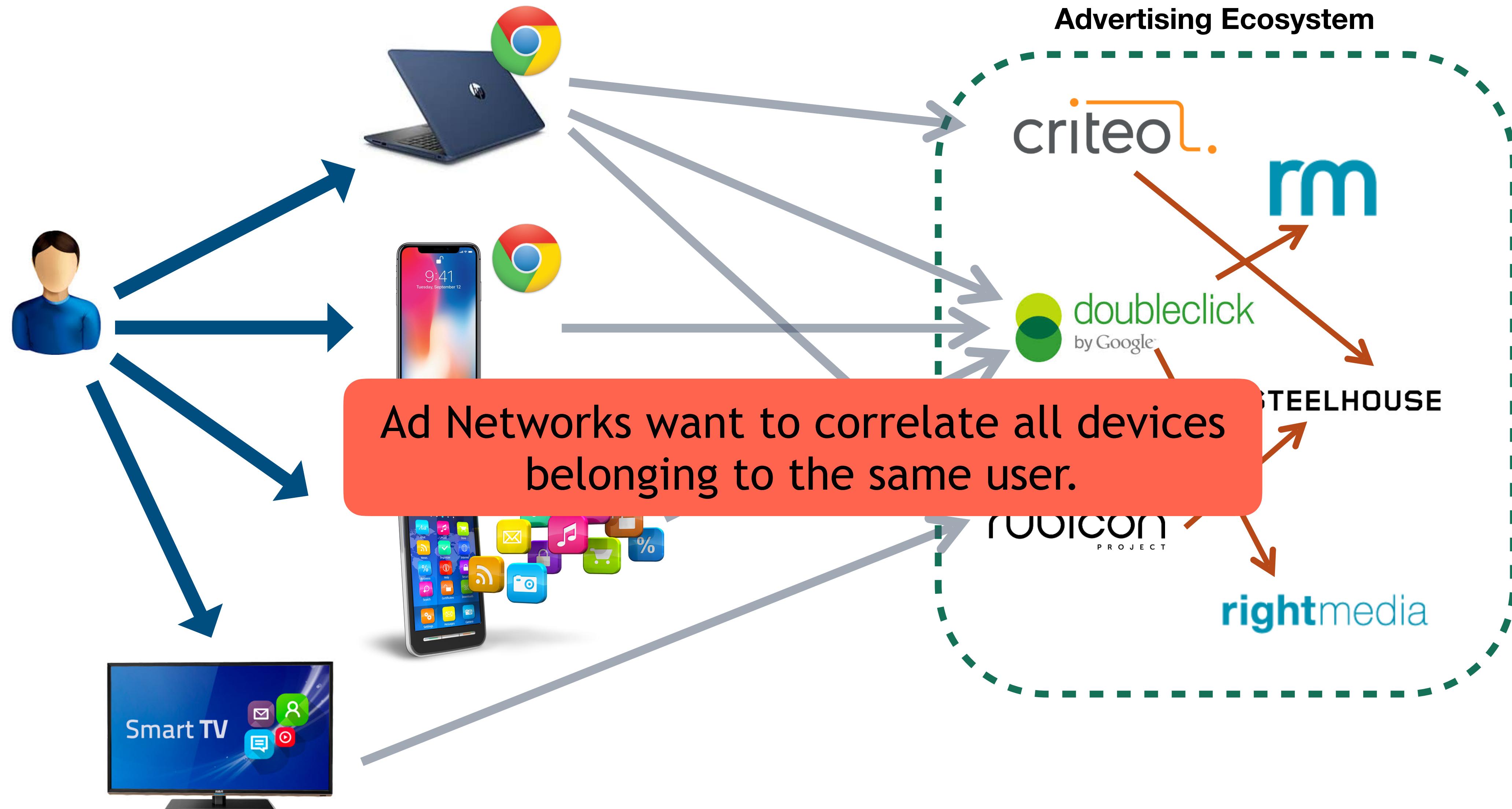
Cross Device Tracking



Cross Device Tracking



Cross Device Tracking



Cross Device Tracking

Key Questions

1. What is the coverage of different tracking companies across devices?
2. How and with whom do these companies share tracking data?
3. Which identifiers give users away across devices?
e.g. email, username, IP address

Challenges

1. Not trivial to instrument mobile devices to collect inclusion chains.
2. Careful setup of controlled experiments needed.

Protection Against Online Tracking

People have grown increasingly concerned about the state of online tracking

Protection Against Online Tracking

People have grown increasingly concerned about the state of online tracking



Protection Against Online Tracking

People have grown increasingly concerned about the state of online tracking



Privacy Badger blocks spying ads
and invisible trackers.

INSTALL PRIVACY BADGER

Click here for Chrome version



GHOSTERY



```
11points.com/images/slack100.jpg
1320wils.com/assets/images/promo%20banner/
1337x.to/js/script.js
1340wcmi.com/images/banners/
1430wnav.com/images/300-
1430wnav.com/images/468-
1590wcgo.com/images/banners/
174.143.241.129^$domain=astalavista.com
1776coalition.com/wp-content/plugins/sam-images/
178.209.48.7^$domain=zerohedge.com
17track.net^*/google.
180upload.com/p1.js
180upload.com/pir/729.js
194.14.0.39/pia.png$domain=tokyo-tosho.net|tokyotosho.info|tokyotosho.se
194.14.0.39/pia_wide.png$domain=tokyo-tosho.net|tokyotosho.info|tokyotosho.se
1movies.to/site/videoroller
1up.com/scripts/takeover.js
1up.com/vip/vip_games.html
1up.com^*/promos/
209.62.111.179/ad.aspx$domain=pachanyc.com
212.7.200.164^$domain=wjunction.com
216.151.186.5^*/serve.php?$domain=sendspace.com
22lottery.com/images/lm468
```

Use Filter lists (e.g. EasyList)

Protection Against Online Tracking

People have grown increasingly concerned about the state of online tracking



Privacy Badger blocks spying ads
and invisible trackers.

INSTALL PRIVACY BADGER

Click here for Chrome version



GHOSTERY

```
11points.com/images/slack100.jpg
1320wils.com/assets/images/promo%20banner/
1337x.to/js/script.js
1340wcmi.com/images/banners/
1430wnav.com/images/300-
1430wnav.com/images/468-
1590wcgo.com/images/banners/
174.143.241.129^$domain=astalavista.com
1776coalition.com/wp-content/plugins/sam-images/
178.209.48.7^$domain=zerohedge.com
17track.net^*/google.
180upload.com/p1.js
180upload.com/pir/729.js
194.14.0.39/pia.png$domain=tokyo-tosho.net|tokyotosho.info|tokyotosho.se
194.14.0.39/pia_wide.png$domain=tokyo-tosho.net|tokyotosho.info|tokyotosho.se
1movies.to/site/videoroller
1up.com/scripts/takeover.js
1up.com/vip/vip_games.html
1up.com^*/promos/
209.62.111.179/ad.aspx$domain=pachanyc.com
212.7.200.164^$domain=wjunction.com
216.151.186.5^*/serve.php?$domain=sendspace.com
22lottery.com/images/lm468
```

Use Filter lists (e.g. EasyList)

- These lists are manually curated over time.
 - Difficult to maintain
 - Whack-a-mole approach

Towards Automation of Adblocking Filter List

Can we come up with a way to automate these filter lists?

- ML based approach from Brave
- An alternate approach based on ADS.txt protocol (Authorized Digital Sellers)

Programmatic Buying & Selling via ADS.txt

Authorized Digital Sellers (ADS.txt)

- Introduced by Interactive Advertising Bureau (IAB) in 2017
- Enables publishers to list authorized sellers

Programmatic Buying & Selling via ADS.txt

Authorized Digital Sellers (ADS.txt)

- Introduced by Interactive Advertising Bureau (IAB) in 2017
- Enables publishers to list authorized sellers

```
# Ads.txt file for washingtonpost.com:  
  
indexexchange.com, 183960, DIRECT  
adtech.com, 10316, DIRECT  
aolcloud.net, 10316, DIRECT  
appnexus.com, 7466, DIRECT  
google.com, pub-3980300725513096, DIRECT  
c.amazon-adsystem.com, 3041, DIRECT
```

Using ADS.txt to Create a Filter List

Idea: We can use ads.txt provided by the publisher

- Crawl the ads.txt file
- Crawl the publisher and collect inclusion chains
- Combine all inclusion chains to form a graph
- Label propagation to detect nodes which should be on filter list

Key Questions

1. Is the coverage good enough?
 - How about exchanges and DSPs coverage?
 - Can we detect trackers as well?
2. How does it compare to current filter lists?

Outline

Background & Motivation

Limitations of Prior Work

Detecting Information Sharing

- Using Retargeted Ads

Propagation of Tracking Data

- Modeling via Graph

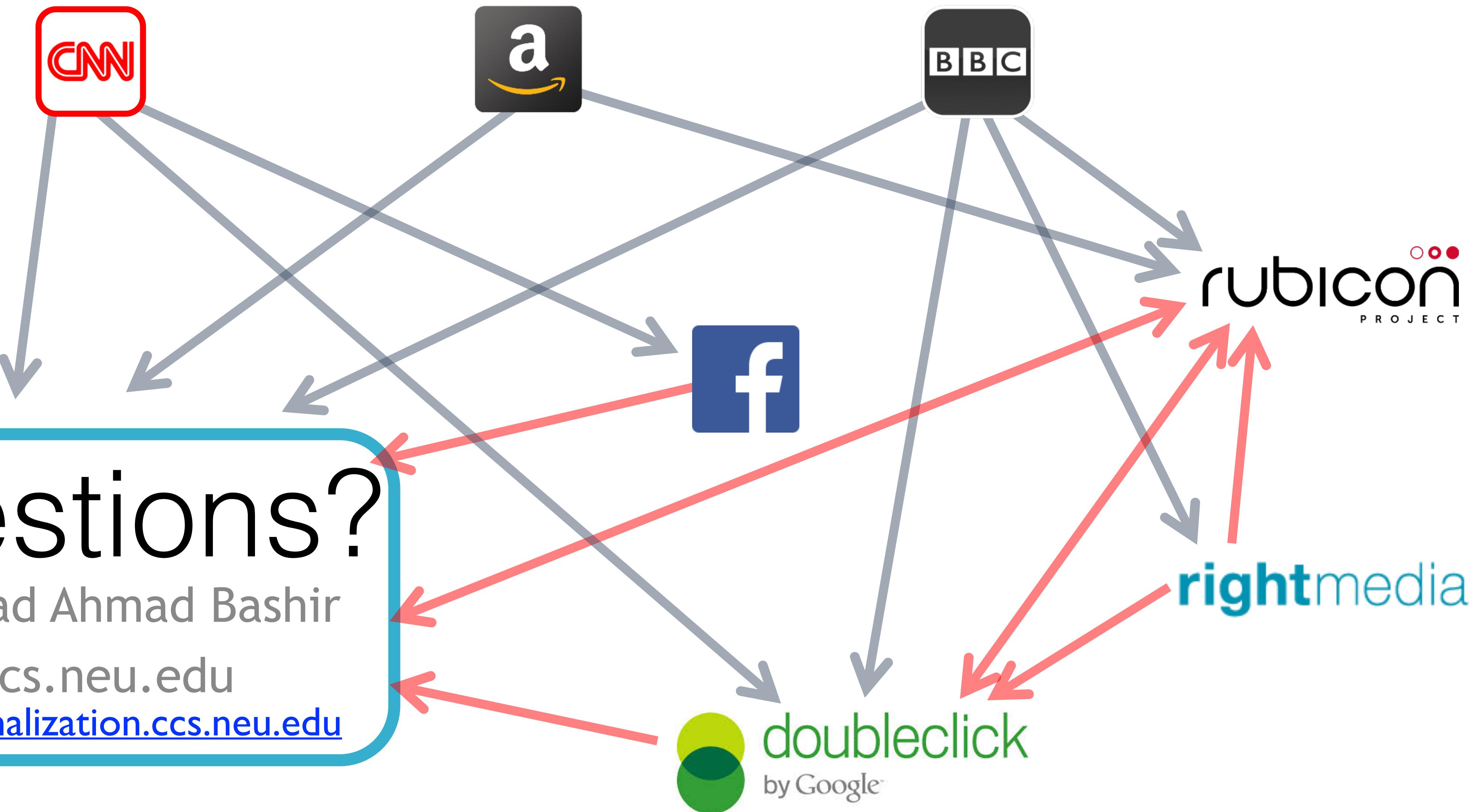
Ongoing Work

- Cross-device Tracking
- Automation of Adblocking Filter Lists

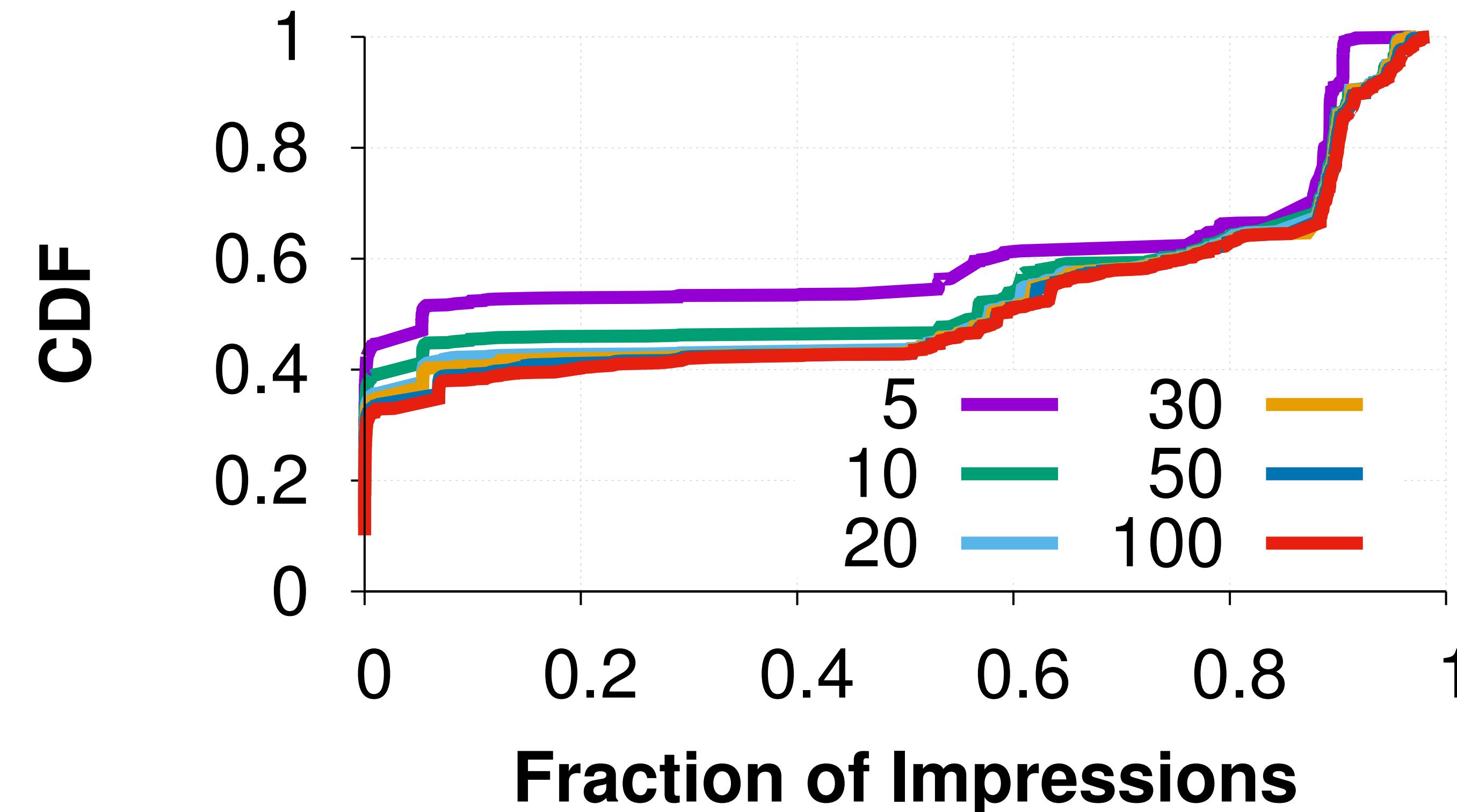
Research Plan

Research Completion Plan

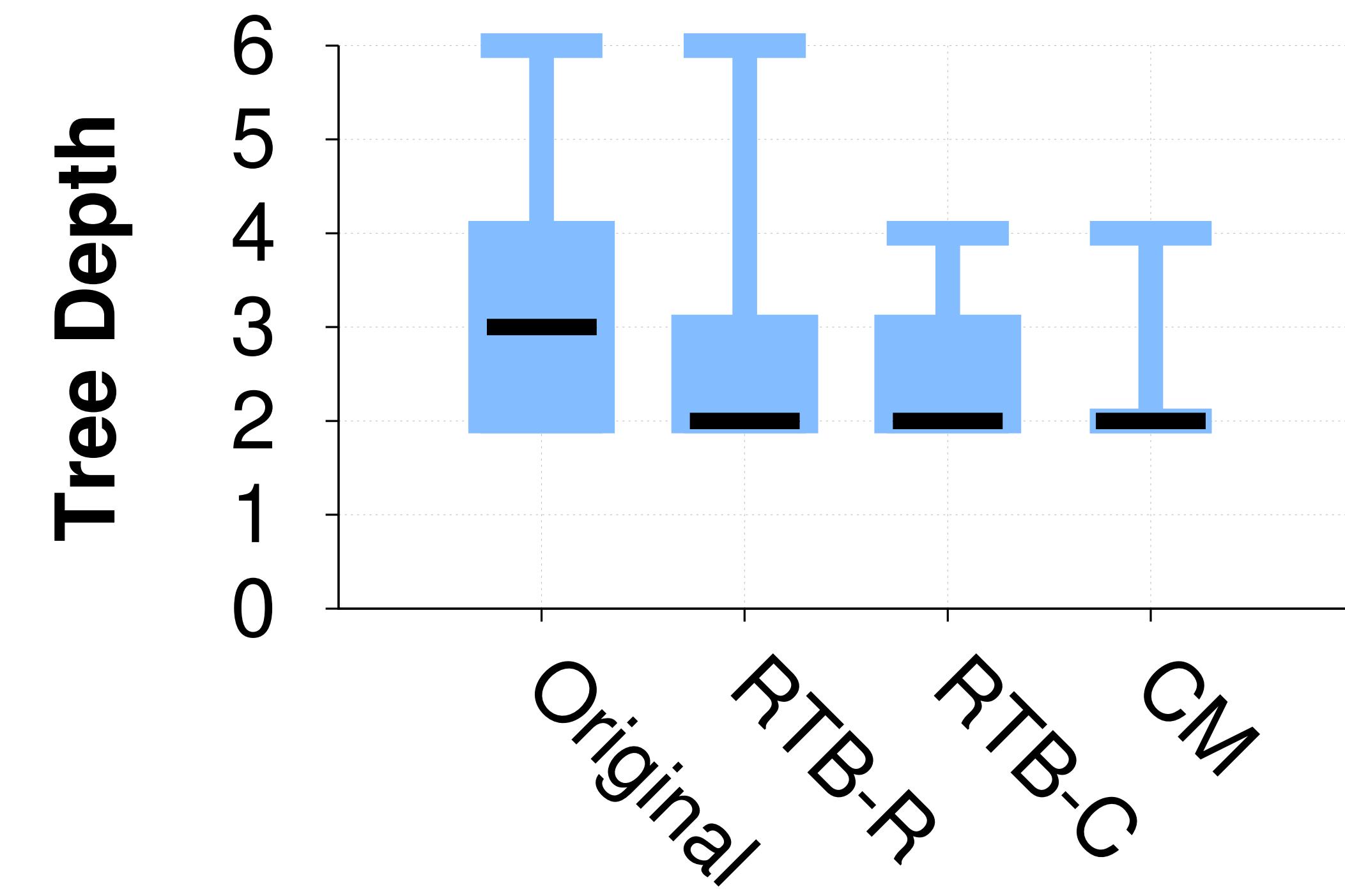
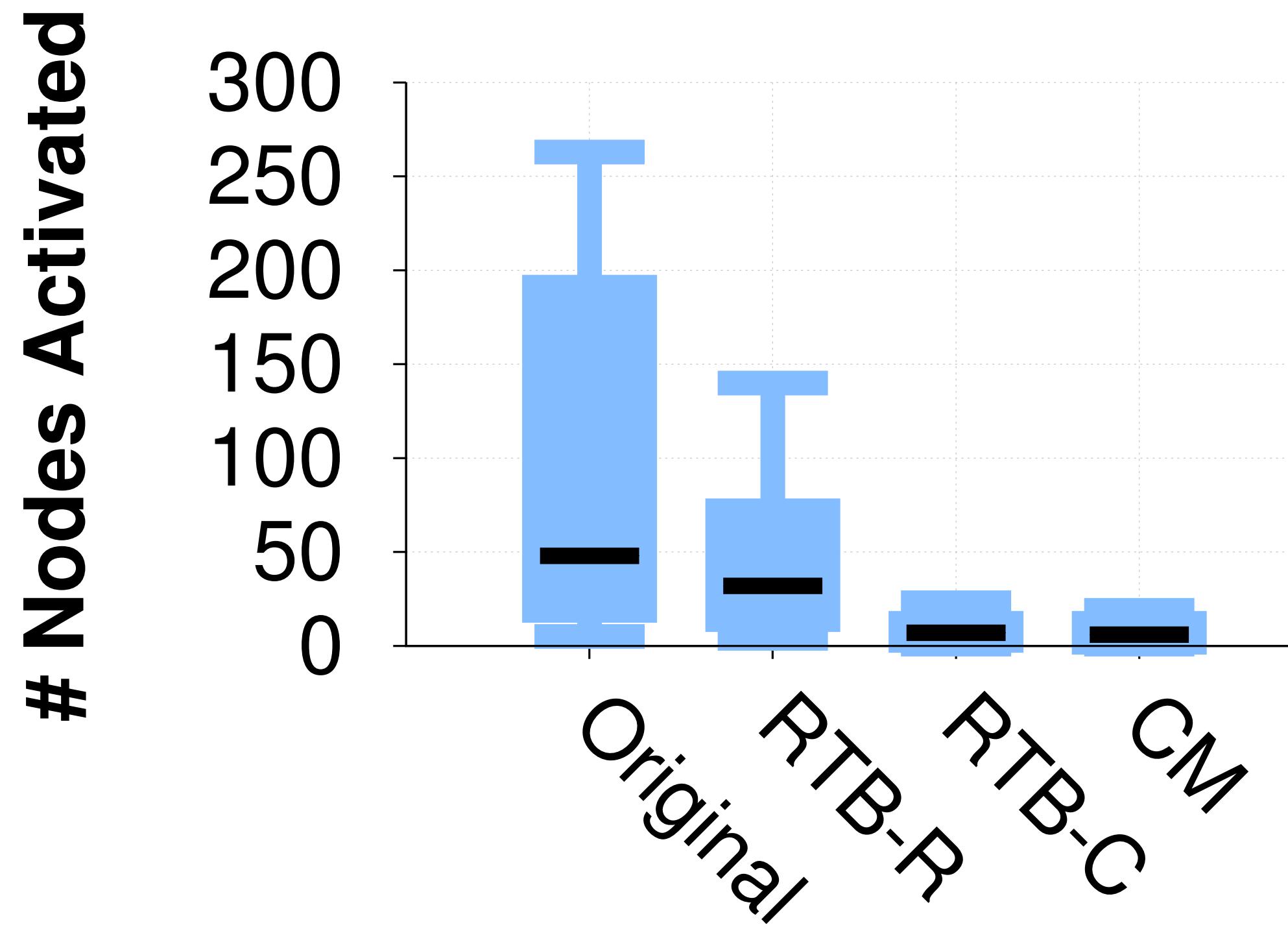
Timeline	Work	Progress
-	Tracing Information Flows Between Ad Exchanges Using Retargeted Ads	Published (Usenix Security '16)
-	Diffusion of User Tracking Data in the Advertising Ecosystem	Published (PETS '18)
Present - APR 2019	Automation of Privacy Protection Filter List	Ongoing
Present - MAY 2019	Cross Device Tracking	Ongoing
MAY - AUG 2019	Thesis Writing	-
AUG 2019	Thesis defense	-



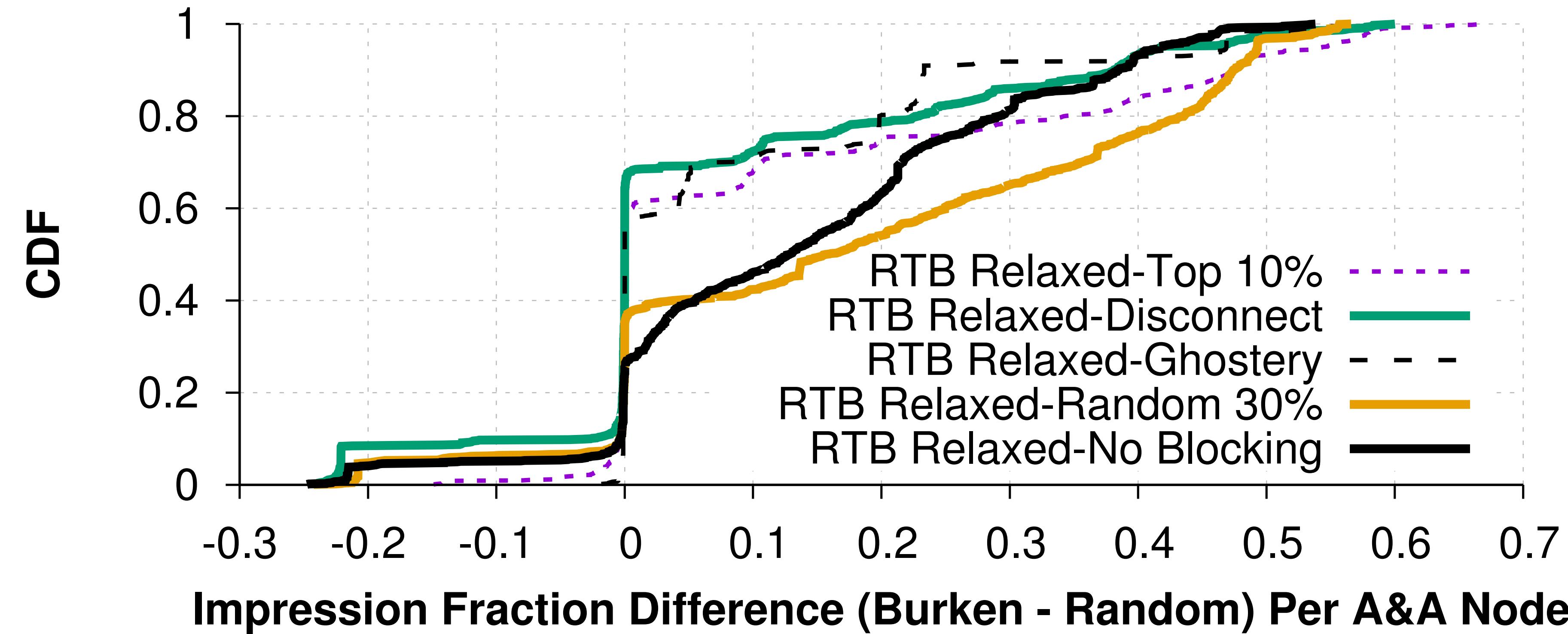
Number of Exchanges (RTB Constrained)



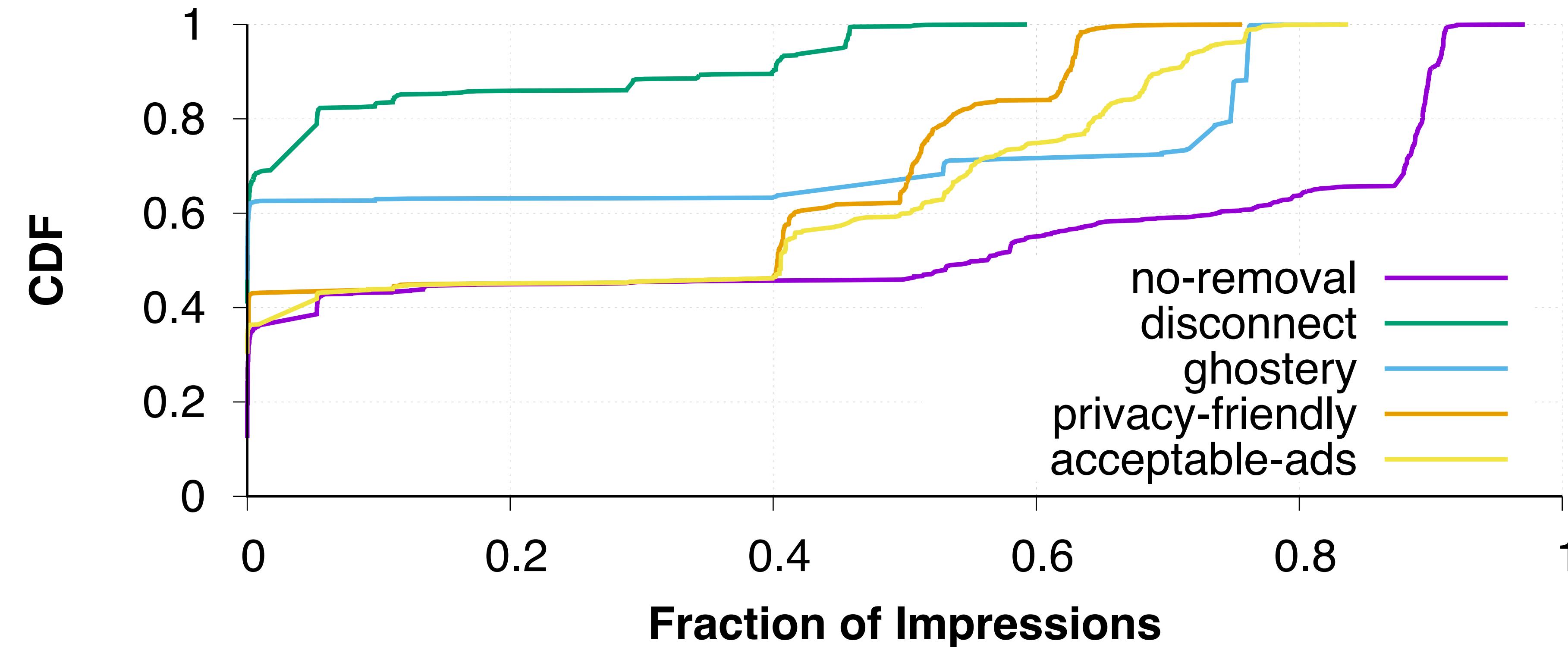
Simulations - Model Validation



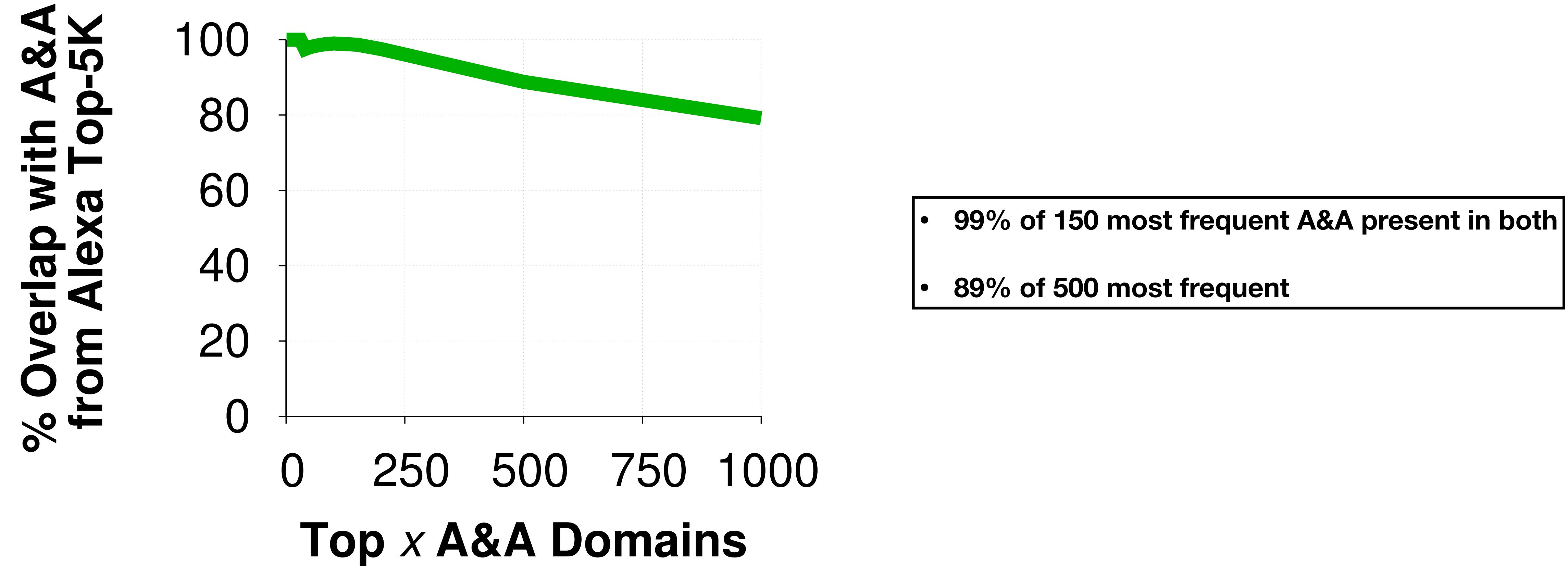
Simulations - Random Model



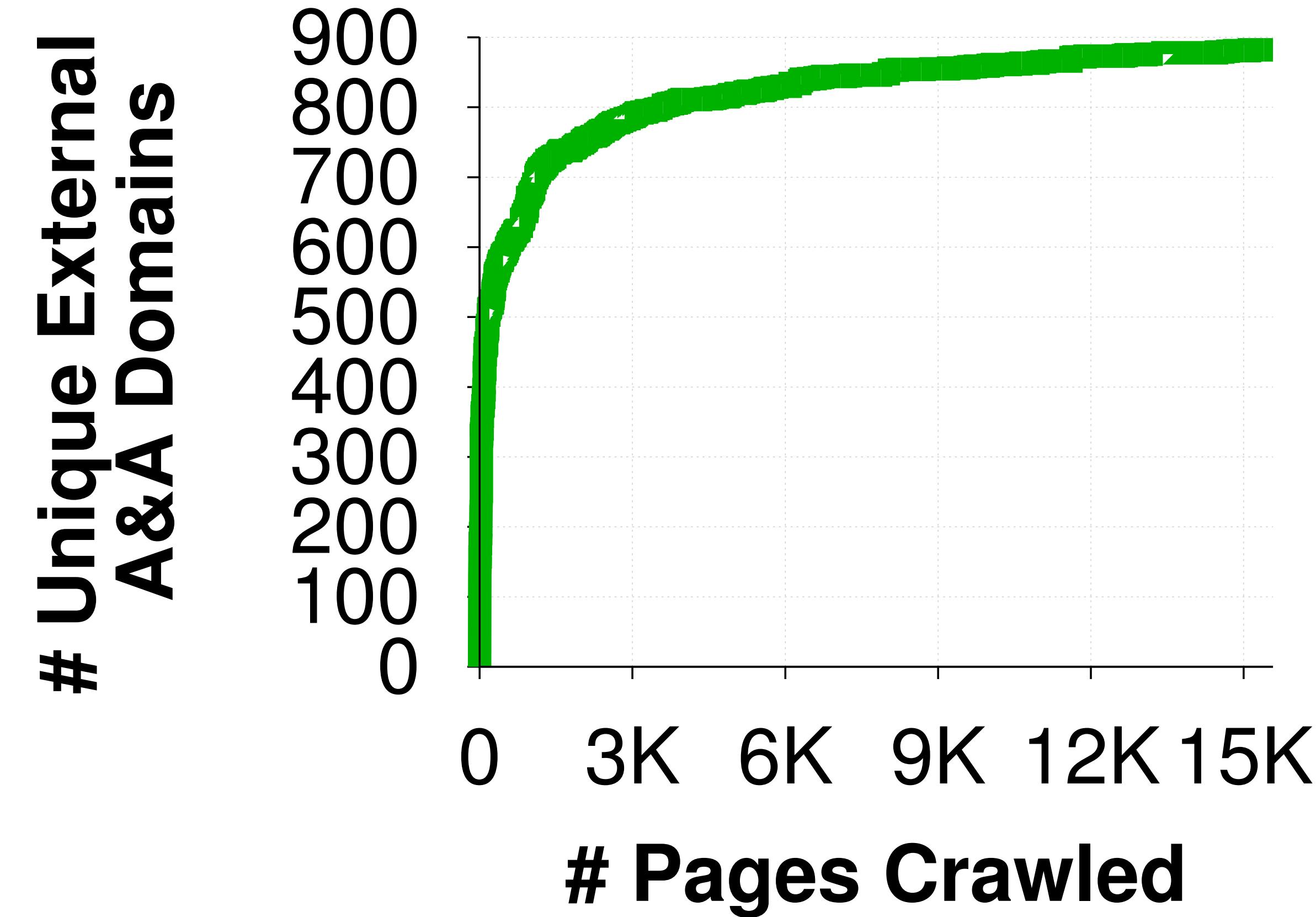
Simulation with Blocking (ABP improved)



A&A Coverage

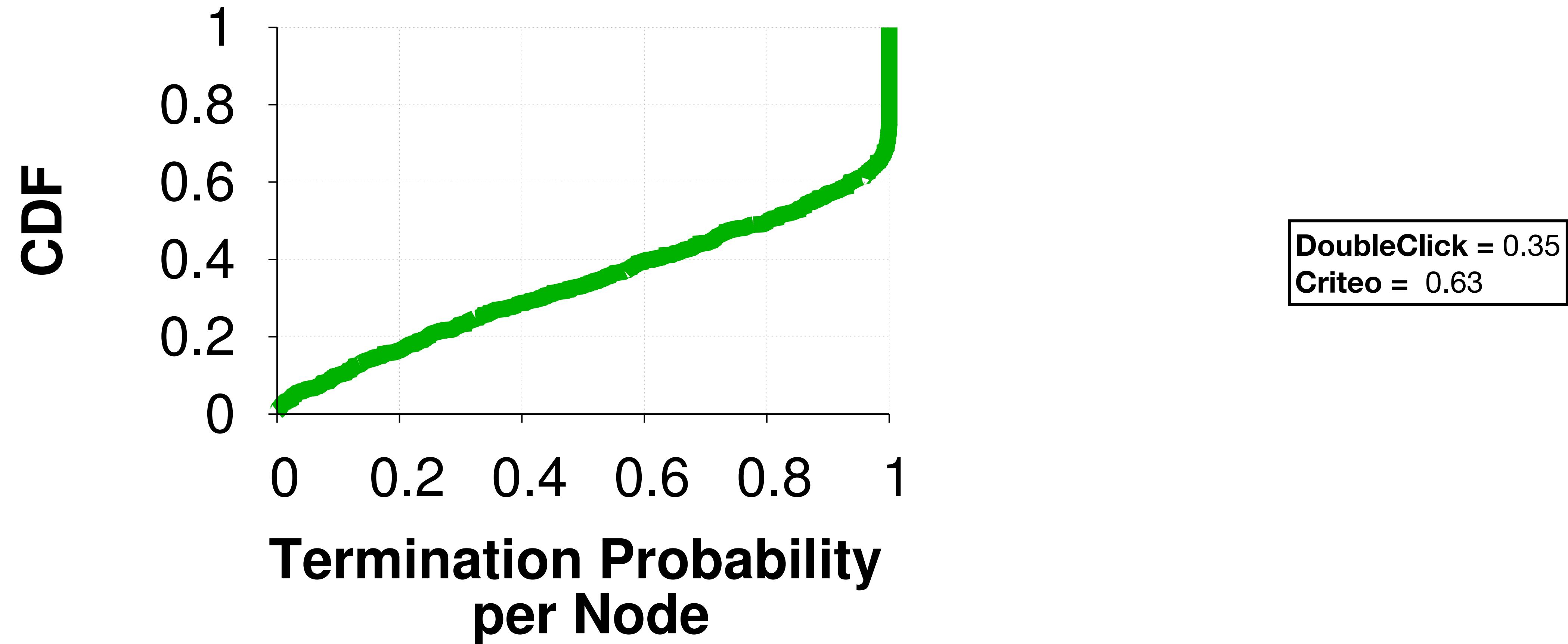


A&A Coverage (2)

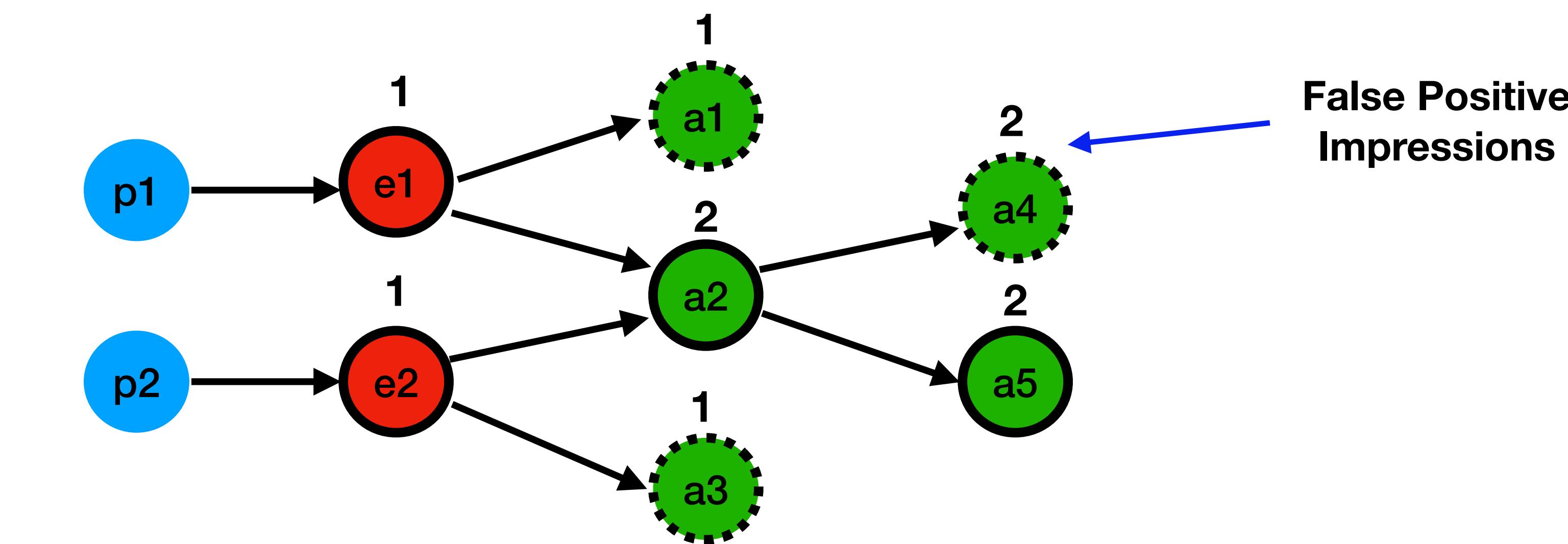
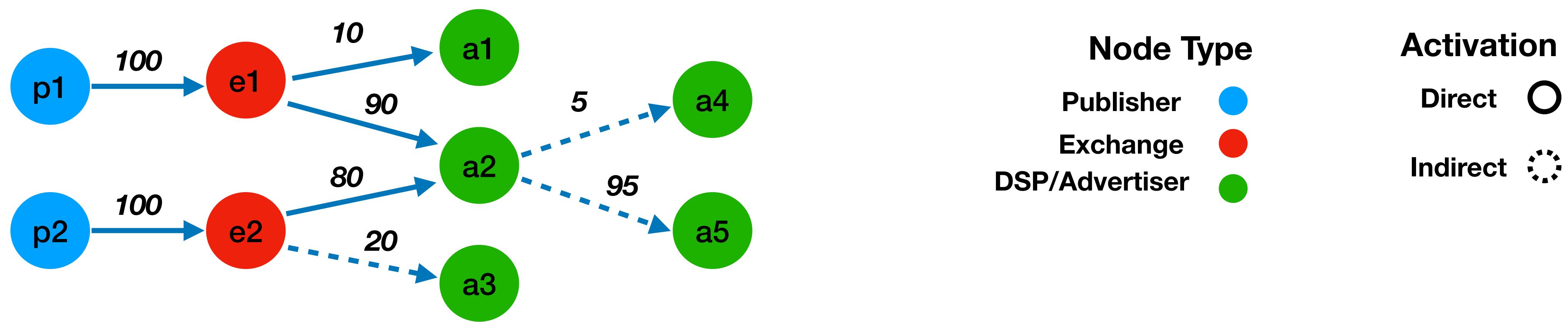


Number of unique A&A contacted by A&A domains as our crawl progresses

Termination Probability



Simulation RTB Relaxed



Simulation Cookie Matching Only

